(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
5 April 2001 (05.04.2001)

PCT

(10) International Publication Number
WO 01/23548 A1

(51) International Patent Classification[7]: C12N 15/11, 9/00, 9/22, C12Q 1/68

(21) International Application Number: PCT/US00/26864

(22) International Filing Date:
29 September 2000 (29.09.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/156,790 30 September 1999 (30.09.1999) US

(71) Applicant: YALE UNIVERSITY [US/US]; 451 College Street, New Haven, CT 06511 (US).

(72) Inventors: DORIT, Robert, L.; 172 Chesterfield Road, Leeds, MA 01053 (US). COLE, Kyle, B.; Apartment 1, 35 Eld Street, New Haven, CT 06511 (US).

(74) Agents: PABST, Patrea, L. et al.; Arnall Golden & Gregory, LLP, 2800 One Atlantic Center, 1201 West Peachtree Street, Atlanta, GA 30309-3450 (US).

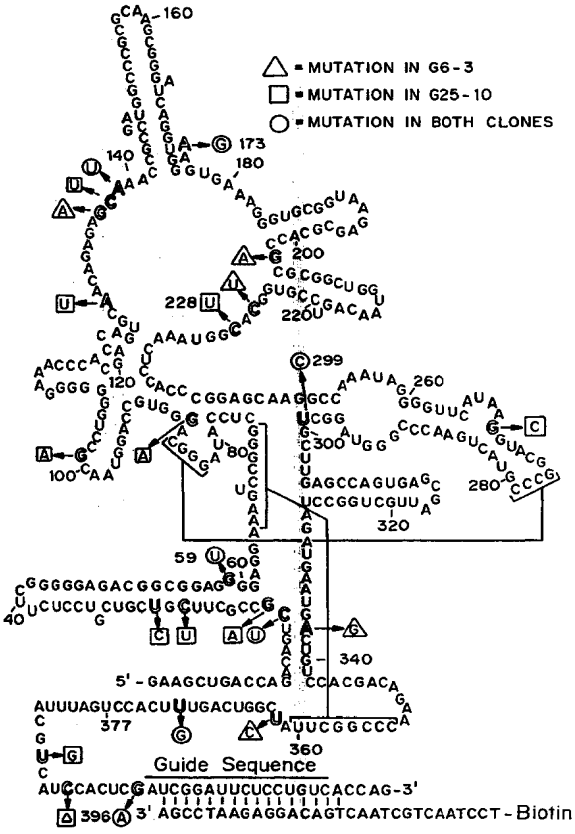
(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *With international search report.*

[Continued on next page]

(54) Title: DNA-CLEAVING RNASE P RNA



(57) Abstract: Disclosed are modified, or variant, RNase P RNA molecules that specifically cleave DNA with a catalytic efficiency greater than wild-type RNase P RNA. Sequence-specific targeting of the variant RNase P RNA is provided by guide sequences that specifically hybridize to the target DNA and cause the variant RNase P RNA to recognize the DNA molecule as a substrate. Thus, the disclosed variant RNase P RNA molecules are universal DNA cleavage enzymes. The disclosed variant RNase P RNA molecules have a wide variety of uses. For example, the disclosed variant RNase P RNA molecules can be used to cleave any desired DNA sequence *in vitro* or *in vivo*. These uses specifically include diagnostic methods for detection, quantitation, or cataloging of DNA sequences, forensic methods, genome dissection methods, biostatistical methods, and population genetics methods, cleaving genomic DNA at particular sequences, creating gene knockouts by gene cleavage, killing specific cells by specific cleavage of DNA, cleaving pathogen DNA in a host cell, and killing mutant cells by specific cleavage of mutant DNA in the cell. Also disclosed is a method of producing variant RNase P RNA molecules that catalyze efficient cleavage of DNA substrates. The disclosed methods allow efficient and predictable generation of a variety of RNase P RNA variants having significant DNA-cleaving activity. The disclosed method thus provides a ready means of producing numerous variant RNase P RNA molecules for use in any of the disclosed methods for use of such variant RNase P RNA molecules. The disclosed variant RNase P RNA molecules can be used with a separate guide sequence molecule or with a guide sequence covalently coupled to the variant RNase P RNA.

WO 01/23548 A1

— *Before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

DNA-CLEAVING RNASE P RNA

BACKGROUND OF THE INVENTION

RNase P is an ubiquitous ribonucleoprotein responsible for processing the 5' termini of several cellular RNAs involved in protein biosynthesis (Altman et al., *Faseb J* 7(1), 7-14 (1993); Brown & Pace, *Biochimie* 73(6), 689-97 (1991); Kirsebom, *Mol. Microbiol.* 17:411-420 (1995)). In *Escherichia coli* this ribonucleoprotein is composed of a 377 nucleotide catalytic RNA moiety, M1 RNA, coupled with a small (14 kD) protein subunit, termed C5 (Stark et al., *Proc. Nat. Acad. Sci. USA* 75:3719-3721 (1977)). M1 RNA retains its catalytic function in vitro in the absence of the C5 protein and – unique among the naturally occurring catalytic RNAs – behaves as a true enzyme, catalyzing a specific reaction without any permanent modification to the ribozyme itself (Guerrier-Takada, *Cell* 35(3 Pt 2):849-57 (1983); Guerrier-Takada et al., *Cell* 38(1):219-224 (1984)).

RNase P processes a remarkable variety of substrates *in vivo*. In *E. coli*, these include all pre-tRNAs, as well as pre-4.5S RNA and 10Sa RNA (Bothwell et al., *J Biol Chem* 251(23), 7709-16 (1976); Hartmann et al., *Proc Natl Acad Sci USA* 92:5822-5826 (1995); Kirsebom & Vioque, *Mol Biol Rep* 22(2-3):99-109 (1995); Komine et al., *PNAS* 91:9223-9227 (1996); Peck-Miller & Altman, *J Mol Biol* 221(1):1-5 (1991)). The structural basis for the recognition of these diverse targets by RNase P has been the subject of detailed investigation and depends on both substrate and ribonucleoprotein features. Essential substrate features include a paired stem ranging from 5 bp to more than 15 bp and at least one overhanging nucleotide in the 5' leader sequence (Altman & Kirsebom, Ribonuclease P, In *The RNA World*, Second Edition, pp. 351-380, Cold Spring Harbor Laboratory Press (1999); Kirsebom & Vioque, *Mol Biol Rep* 22(2-3):99-109 (1995); Liu & Altman, *Nucleic Acids Res* 24(14):2690-2696 (1996)). The majority of substrates also include a 3'-RCCA motif, required by M1 RNA *in vitro* for efficient processing in the absence of C5 protein. RNase P versatility also depends on the ability of the ribonucleoprotein to form tertiary contacts with a variety of substrates. These contacts are mediated, at least in part, by the C5 protein

(Kirsebom & Vioque, *Mol Biol Rep* 22(2-3):99-109 (1995); Vioque et al., *J Mol Biol* 202(4):835-48 (1988)).

Attempts to understand the mechanism of M1 RNA catalysts have spurred extensive structural and biochemical characterization of this universal component of metabolism (Chen et al., *Embo J* 17(5), 1515-1525 (1998); Ciesiolka et al., *Eur J Biochem* 219(1-2), 49-56 (1994); Harris et al., *EMBO* 13(17):3953-3963 (1994); Harris et al., *RNA* 3(6):561-576 (1997); Harris & Pace, *RNA* 1(2):210-218 (1995); LaGrandeur et al., *Embo J* 13(17):3945-3952 (1994); Lawrence & Altman, *J. Mol Biol* 191(2):163-75 (1986); Lumelsky & Altman, *J Mol Biol* 202(3):443-54 (1988); Westhof et al., *J. Mol. Biol.* 258(4):600-613 (1996)). More recently, experimental *in vitro* selection techniques have supplemented the conventional arsenal of tools for the study of molecular structure and function. These *in vitro* methods provide a kind of perturbation analysis, where the wild-type sequence can be mutagenized (or partially randomized), and the resulting variants screened for active molecules. Such approaches have been applied to M1 RNA and its substrates to explore the structural and catalytic features of this ribozyme. *In vitro* selection has produced M1 RNA derivatives with altered divalent metal ion specificity (Frank & Pace, *Proc Natl Acad Sci U S A* 94(26):14355-14360 (1997)) and has been used to support the assertion that the catalytic core of this ribozyme is evolutionarily optimized (Frank et al., *RNA* 2:1179-1188 (1996); Kim et al., *RNA* 3(6):613-623 (1997)) Alternate RNA substrates for RNase P have also been produced by *in vitro* selection (Liu & Altman, *Cell* 77(7):1093-1100 (1994); Pan, *Biochemistry* 34(26):8458-8464 (1995); Pan & Jakacka, *Embo J.* 15(9):2249-2255 (1996)). These studies have expanded the range of artificial substrates and, in so doing, highlighted the structural features that define a substrate for this ribozyme.

Objects of the Invention

It is therefore an object of the present invention to provide variant RNase P RNA molecules that catalyze efficient cleavage of DNA substrates.

It is another object of the present invention to provide a method of producing variant RNase P RNA molecules that catalyze efficient cleavage of DNA substrates.

It is another object of the present invention to provide variant RNase P RNA molecules that catalyze efficient cleavage of any desired target DNA molecule.

It is another object of the present invention to provide a method of cleaving any desired DNA molecule using variant RNase P RNA molecules that catalyze efficient cleavage of DNA substrates.

It is another object of the present invention to provide diagnostic methods for detection, quantitation, or cataloging of DNA sequences using variant RNase P RNA molecules that catalyze efficient cleavage of DNA substrates.

It is another object of the present invention to provide use of variant RNase P RNA molecules to cleave genomic DNA at particular sequences.

It is another object of the present invention to provide use of variant RNase P RNA molecules that catalyze efficient cleavage of DNA substrates in forensic methods, genome dissection methods, biostatistical methods, and population genetics methods.

It is another object of the present invention to provide a method of creating gene knockouts by targeting specific genes for cleavage by RNase P RNA molecules that catalyze efficient cleavage of DNA substrates.

It is another object of the present invention to provide a method of killing specific cells by specific cleavage of DNA using variant RNase P RNA molecules that catalyze efficient cleavage of DNA substrates.

It is another object of the present invention to provide a method of cleaving pathogen DNA in a host cell using variant RNase P RNA molecules that catalyze efficient cleavage of DNA substrates.

It is another object of the present invention to provide a method of killing mutant cells by specific cleavage of mutant DNA in the cell using variant RNase P RNA molecules that catalyze efficient cleavage of DNA substrates.

It is another object of the present invention to provide variant RNase P RNA molecules that include a guide sequence and that catalyze efficient cleavage of DNA substrates.

BRIEF SUMMARY OF THE INVENTION

Disclosed are modified, or variant, RNase P RNA molecules that specifically cleave DNA with a catalytic efficiency greater than wild-type RNase P RNA. The disclosed variant RNase P RNA can cleave any desired DNA sequence since targeting of the RNase P RNA and efficient DNA catalysis are substantially independent. Sequence-specific targeting of the variant RNase P RNA is preferably provided by guide sequences that specifically hybridize to the target DNA and cause the variant RNase P RNA to recognize the targeted DNA molecule as a substrate. Thus, the disclosed variant RNase P RNA molecules are universal DNA cleavage enzymes.

The disclosed variant RNase P RNA molecules have a wide variety of uses. For example, the disclosed variant RNase P RNA molecules can be used to cleave any desired DNA sequence *in vitro* or *in vivo*. These uses specifically include diagnostic methods for detection, quantitation, or cataloging of DNA sequences, forensic methods, genome dissection methods, biostatistical methods, and population genetics methods, cleaving genomic DNA at particular sequences (in preparation for use in cloning, diagnostic, or detection methods, for example), creating gene knockouts by gene cleavage, killing specific cells by specific cleavage of DNA, cleaving pathogen DNA in a host cell, and killing mutant cells by specific cleavage of mutant DNA in the cell.

Use of the disclosed variant RNase P RNA molecules for preventing or altering expression of a gene *in vivo* has advantages over most other approaches to such regulation (for example, antisense RNA and transcription inhibitors) since the disclosed molecules target and inactivate the gene. In this way, the cell cannot circumvent regulation by increasing the number of transcripts of the gene as is the potential with other forms of regulation. The disclosed molecules have the additional advantage in that they can attack dormant (that is, non-expressed) genes while many other forms of regulation

require gene expression to be effective. Thus, the disclosed molecules can be particularly effective in attacking dormant viruses in cells.

Also disclosed is a method of producing variant RNase P RNA molecules that catalyze efficient cleavage of DNA substrates. The disclosed method allows efficient and predictable generation of a variety of RNase P RNA variants having significant DNA-cleaving activity. The disclosed method thus provides a ready means of producing numerous variant RNase P RNA molecules for use in any of the disclosed methods for use of such variant RNase P RNA molecules. The disclosed variant RNase P RNA molecules can be used with a separate guide sequence molecule or with a guide sequence covalently coupled to the variant RNase P RNA. The disclosed variant RNase P RNA molecules can be used without a guide sequence if the variant was selected to recognize a target molecule directly. The disclosed variant RNase P RNA molecules can be used with an RNase P protein subunit. An RNase P protein subunit is preferably used if the variant was selected in the presence of an RNase protein subunit.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1A is a diagram of the secondary structure of an example of an RNase P RNA construct (in this case, M1GS) including a covalently-tethered 3' guide sequence complementary to the DNA oligonucleotide target substrate. The RNase P construct is SEQ ID NO:1. The target substrate is SEQ ID NO:3.

Figure 1B is diagram of an example of a selection scheme used to produce DNA-cleaving RNase P variants. (1) M1GS variants are annealed to a 5'-biotinylated DNA substrate in a buffer that allows Watson-Crick base-pairing but prevents cleavage; (2) the M1GS/substrate complex is bound to streptavidin-coated paramagnetic beads and unbound ribozymes are removed by extensive washing; (3) upon addition of a buffer containing Mg^{2+} ions, ribozymes capable of performing catalysis under selection conditions are released from the beads; (4) active ribozymes are retrieved

from the supernatant, reverse transcribed and amplified through standard or mutagenic PCR.

Figure 2 is a graph of the location and frequency of mutations found in G6 and G25 populations produced during the selection process. Frequency is based on sampling forty sequences from each population. Note that three of the four fixed mutations found in G25 (59:U or C, 173:G, 228:U and 299:C) occur in at least 30% of the G6 population sequences. Cross-hatched boxes indicate ribozyme domains where a significant number of mutations accumulate. The P4 pseudoknot (stippled boxes), the central component of the catalytic core, does not contain mutations.

Figure 3 is a diagram of the secondary structure of RNase P RNA (SEQ ID NO:1) with specific mutations occurring in representative clones drawn from generation six (G6-3) and from generation twenty-five (G25-10) superimposed. The target substrate is SEQ ID NO:3

Figure 4A is a diagram of substrate configurations for a *cis* cleavage reaction by RNase P RNA.

Figure 4B is a diagram of substrate configurations for several *trans* cleavage reactions by RNase P RNA.

Figure 5 is a graph of the relative efficiency (k_{cat}/K_m) of wild-type RNase P RNA and representative evolved RNase P RNA molecules (G6-3 and G25-10) on various DNA and RNA substrates.

Figure 6 is a diagram of the sequence and proposed secondary structure of H1 RNA (SEQ ID NO:2), the RNA component of human RNase P.

Figure 7 is a diagram of the sequence and proposed secondary structure of M1 RNA (nucleotides 1 to 377 of SEQ ID NO:1), the RNA component of *E. coli* RNase P.

Figure 8 is a diagram of an example of a selection scheme used to produce DNA-cleaving RNase P variants. (A) M1K1 variants are annealed to a 5'-biotinylated DNA substrate in a buffer that allows Watson-Crick base-pairing but prevents cleavage. (B) The M1GS/substrate complex is bound to streptavidin-coated paramagnetic beads and unbound ribozymes are removed

by extensive washing. (C) Upon addition of a buffer containing Mg2+ ions, ribozymes capable of performing catalysis under selection conditionsare released from the beads and retrieved from the supernatant. (D) Subsequent reverse-transcription, PCR amplification, and transcription create the next ribozyme population.

Figure 9 is a graph of DNA cleavage efficiencies (expressed as log k_{cat}/K_M ($min^{-1}M^{-1}$)) for M1K1 RNase P RNA (using the wild type sequence), G10.5 variant RNase P RNA, and both with various mutations.

DETAILED DESCRIPTION OF THE INVENTION

Disclosed are modified, or variant, RNase P RNA molecules that specifically cleave DNA with a catalytic efficiency greater than wild-type RNase P RNA. Sequence-specific targeting of the variant RNase P RNA is provided by guide sequences that specifically hybridize to the target DNA and cause the variant RNase P RNA to recognize the DNA molecule as a substrate. Thus, the disclosed variant RNase P RNA molecules are universal DNA cleavage enzymes.

The disclosed variant RNase P RNA molecules have a wide variety of uses. For example, the disclosed variant RNase P RNA molecules can be used to cleave any desired DNA sequence *in vitro* or *in vivo*. These uses specifically include diagnostic methods for detection, quantitation, or cataloging of DNA sequences, forensic methods, genome dissection methods, biostatistical methods, and population genetics methods, cleaving genomic DNA at particular sequences, creating gene knockouts by gene cleavage, killing specific cells by specific cleavage of DNA, cleaving pathogen DNA in a host cell, and killing mutant cells by specific cleavage of mutant DNA in the cell.

Also disclosed is a method of producing variant RNase P RNA molecules that catalyze efficient cleavage of DNA substrates. The disclosed methods allow efficient and predictable generation of a variety of RNase P RNA variants having significant DNA-cleaving activity. The disclosed method thus provides a ready means of producing numerous variant RNase P

RNA molecules for use in any of the disclosed methods for use of such variant RNase P RNA molecules. The disclosed variant RNase P RNA molecules can be used with a separate guide sequence molecule or with a guide sequence covalently coupled to the variant RNase P RNA. The disclosed variant RNase P RNA molecules can be used without a guide sequence if the variant was selected to recognize a target molecule directly. The disclosed variant RNase P RNA molecules can be used with an RNase P protein subunit. An RNase P protein subunit is preferably used if the variant was selected in the presence of an RNase protein subunit.

The ribonucleoprotein RNase P is a critical component of metabolism in all known organisms. In *Escherichia coli*, RNase P processes a vast array of substrates, including precursor-tRNAs and precursor-4.55 RNA. The M1 RNA ribozyme—the catalytic component of *E. coli* RNase P—was evolved *in vitro* for cleavage of a DNA substrate. Twenty-five generations of *in vitro* evolution yielded a population showing a thousand-fold increase in DNA substrate cleavage efficiency (k_{cat}/K_M) relative to wild-type M1 RNA. The enhanced cleavage of a DNA substrate observed in evolved M1 RNA derivatives results from an acceleration of the catalytic step (k_{cat}) of DNA cleavage, although overall processing efficiency was offset by reduced substrate binding. Ten generations of *in vitro* evolution in the presence of the RNase P protein subunit C5 yielded variant RNase P RNAs that cleave DNA more efficiently than wild type RNase P RNA. Thus, use of an RNase P protein subunit during selection increases the efficiency of the in vitro evolution.

I. Compositions

The disclosed variant RNase P RNA molecules are derived from wild-type RNase P RNA. The variant RNase P RNA molecules are used in combination with a guide sequence which may be either a separate molecule or covalently coupled, linked, or tethered to the RNase P RNA. The guide sequence may also be a part of the target DNA to be cleaved. The disclosed variant RNase P RNA molecules can be used without a guide sequence if the variants were selected to recognize a target molecule directly. The disclosed

variant RNase P RNA molecules can be used with RNase P protein subunits. This is especially useful if the variants were selected in the presence of an RNase P protein subunit. The target DNA to be cleaved can be any DNA molecule of interest. The variant RNase P RNA molecules can be used to cleave DNA in numerous contexts. For use in cells, the variant RNase P RNA can be introduced into cells directly or, more preferably, an expression construct encoding the variant RNase P RNA can be introduced into the cell. Expression of the variant RNase P RNA in the cell then allows the variant RNase P RNA to cleave target DNA in the cell. Numerous carriers and delivery compounds and compositions are suitable for use with the disclosed RNase P RNA molecules to deliver the molecules to cells.

The variant RNase P RNA molecules are produced using the disclosed method of *in vitro* selection. For this method, the RNase P RNA is preferably associated with a guide sequence. Alternatively, the RNase P RNA can be selected for cleavage of target molecule directly without the use of a guide sequence. In the selection, a DNA molecule is used as the substrate in order to select for DNA cleavage activity. In a preferred mode of the selection, the substrate DNA is associated with a capture tag. This allows the two halves of the substrate DNA to be separated once cleavage occurs by interaction of the capture tag with a capture tag receptor. This in turn allows separation of RNase P RNA molecules that can cleave DNA efficiently from those that cannot. Once an active variant RNase P RNA molecule is identified, the variant RNase P RNA can be associated with any suitable guide sequence targeted to any desired target DNA sequence. That is, the selected variant RNase P RNA need not be used with the same guide sequence used in the selection process. Variant RNase P RNA molecules selected without the use of a guide sequence need not be associated with a guide sequence. The variant RNase P RNAs can also be selected in the presence of RNase P protein subunit.

A. RNase P RNA

The disclosed molecules, compositions, and methods all make use of variant RNase P RNA that has significant DNA-cleaving activity. Such

variant RNase P RNAs are derived from wild-type RNase P RNA, which is the catalytic subunit of the ribonucleoprotein RNase P.

1. Ribonuclease P

Ribonuclease P (RNase P) is an enzyme consisting of protein and RNA subunits that cleaves tRNA precursors to generate the 5' termini of tRNAs. This essential enzymatic activity has been found in all cell types examined, both prokaryotic and eukaryotic. During the studies on recognition of substrate by RNase P, it was found that *E. coli* RNase P can cleave synthetic tRNA-related substrates that lack certain domains, specifically, the D, T and anticodon stems and loops, of the normal tRNA structure. A half-turn of an RNA helix and a 3' proximal CCA sequence contain sufficient recognition elements to allow the reaction to proceed. The 5' proximal sequence of the RNA helix does not have to be covalently linked to 3' proximal sequence of the helix. The 3' proximal sequence of the stem can be regarded as a "guide sequence" because it identifies the site of cleavage in the 5' proximal region through a base-paired region.

RNase P from *E. coli* and human cells have similar but not identical biochemical properties. Their RNA components have similar secondary structures. However, the substrate range of human RNase P is much narrower than that of the *E. coli* enzyme. For example, although *E. coli* RNase P can cleave a synthetic tRNA-related substrate that lacks three specific domains of the normal tRNA structure, the human enzyme and the structurally similar enzyme from the yeast, *S. cerevisiae*, cannot cleave the same substrate. However, the *E. coli* RNase P can cleave a synthetic tRNA-related substrate that is also cleaved by the human RNase P. Altman *et al.*, *Genomics* 18:419-422 (1993), describes several mammalian RNase P catalytic RNAs and identifies common features and differences. Wild-type ribonuclease P cleaves a DNA substrate inefficiently: 10^2-10^4-fold less efficiently than an analogous RNA substrate (Kleineidam et al., *Nucleic Acids Res* 21(5):1097-1101 (1993); Li & Altman, *Nucleic Acids Res.* 24(5):835-842 (1996); Perreault & Altman, *J Mol Biol* 226(2):399-409 (1992)).

2. Ribonuclease P RNA

Ribonuclease P RNA (RNase P RNA) is both a component used in the disclosed method to produce variant RNase P RNA molecules and the key component of the disclosed methods for cleaving DNA sequences. Any form of RNase P RNA, including previously identified variant RNase P RNAs, can be used as the starting RNase P for selection of variant RNase P RNA having DNA-cleavage activity. The sequence and proposed secondary structure of H1 RNA, the RNA component of human RNase P, was reported by Altman et al., *Faseb J* 7(1), 7-14 (1993), the teachings of which are generally known. The sequence and proposed structure of H1 RNA is shown in Figure 6. The sequence and proposed secondary structure of M1 RNA, the RNA component of *E. coli* RNase P, was reported by James *et al.*, *Cell* 52: 19 (1988), the teachings of which are generally known in the art. The sequence of M1 RNA is shown in Figure 7. The disclosed selection method does not depend on the structure of the RNase P RNA nor knowledge of the structure.

The disclosed variant RNase P RNA molecules are made by alteration of wild-type RNase P RNA (or by further alteration of variant RNase P RNA). For selection of variants, the RNase P RNA is preferably used in conjunction with a guide sequence (GS). The guide sequence is an oligonucleotide that hybridizes to the target DNA sequence, thus forming a structure recognized by RNase P RNA as a substrate for cleavage. A guide sequence may be a separate molecule, termed an external guide sequence, or combined in a single molecule with RNase P RNA. Such a combined molecule is referred to herein as an RNase P internal guide sequence (RIGS). For use of the disclosed variant RNase P RNA molecule to cleave DNA, it is preferred that the guide sequence be combined in a single molecule with RNase P RNA. Thus, the guide sequence can be, for example, covalently coupled, directly or indirectly, to the RNase P, or tethered to the RNase P via a spacer or linker molecule. The disclosed variant RNase P RNA molecules can be used without a guide sequence by selecting the variant to recognize a target molecule directly.

RNase P RNA can be made in any suitable manner including direct synthesis and transcription from a gene, either *in vivo* or *in vitro*. It is preferred that the RNase P RNA be made by transcription. For ultimate use in cells, it is preferred that the RNase P RNA be transcribed in the cell from an expression construct encoding the RNase P RNA introduced into the cell.

As used herein unless otherwise stated, RNase P refers to the ribonucleoprotein consisting of prokaryotic or eukaryotic analogues of the *E. coli* C5 protein and M1 RNA, regardless of source, whether isolated, or produced by chemical synthesis. RNase P RNA refers to the RNA subunit of RNase P, including the disclosed variants. Wild-type RNase P RNA refers to RNase P RNA in any form existing in nature. Variant RNase P RNA refers to a modified form of RNase P RNA that differs from the chemical structure of wild-type RNase P RNA. Most commonly, variant RNase P RNAs differ from wild-type RNase P RNA in the sequence of nucleotides. As used herein, the term variant RNase P RNA encompasses RNase P RNA molecules combined in a single molecule with a guide sequence, although the separate components will usually be specified for the sake of clarity. The term wild-type RNase P RNA does not encompass a covalently coupled guide sequence (since such a configuration does not occur in nature). The eukaryotic RNase P RNA subunit is referred to as H1 RNA. The protein component of RNase P (for example, the C5 protein) is referred to herein as RNase P protein or RNase P protein subunit.

The preferred form of the RNase P RNA may vary depending on in which method it is to be used. In general, it is preferred that the RNase P RNA be covalently coupled to a guide sequence. This simplifies the DNA cleavage reaction, making it a bimolecular reaction rather than a trimolecular reaction if a separate guide sequence is used. Separate RNase P RNA and guide sequence molecules are useful, however, for cleaving multiple target DNA sequences using a single form of RNase P RNA. The disclosed variant RNase P RNA molecules can be used without a guide sequence if the variant was selected to recognize a target molecule directly.

3. Ribonuclease P Protein

The protein component of RNase P (for example, the C5 protein) is referred to herein as RNase P protein or RNase P protein subunit. RNase P protein subunits can be used both during selection of variant RNase P RNAs and during use of the variant RNase P RNAs to cleave DNA or RNA. RNase P protein subunits can be made in any suitable manner including recombinant production, by purification from natural or recombinant sources, and direct synthesis. It is preferred that the RNase P protein subunits be made recombinantly. For ultimate use in cells, it is preferred that the RNase P protein subunits be expressed in the cell. Expression can be from a recombinant construct introduced into the cell or, since cells naturally produce RNase P protein subunits, the endogenous protein subunits may be used.

C5 protein can be isolated following the procedure detailed by Gopalan et al., *Biochemistry* **38**(6):1705-1714 (1999). Briefly, BL21(DE3) cells can be transformed with the plasmid pBSC5 containing wild-type C5. The cells can be grown to an $A_{600} = 0.4$ and then induced with 2 mM IPTG. The crude cell extract is then prepared by sonication and C5 protein concentrated by a series of centrifugations. The protein can be further purified using CM Sephadex C50 chromatography. Protein concentration can be determined using a Bradford assay and measuring the absorbance at 595 nM.

4. Variant RNase P RNA

Disclosed are variant RNase P RNA molecules having DNA-cleavage activity. These include any variant RNase P RNA molecules produced by the disclosed method of selecting variant RNase P RNA molecules having increased DNA-cleavage activity. Preferred variant RNase P RNA molecules include RNase P RNA molecules that contain one or more of the following alterations: a C at position 18, an A at position 19, a U at position 26, a C at position 28, a U at position 29, a U at position 59, an A at position 87, an A at position 101, a U at position 129, a U at position 136, an A at position 137, a U at position 138, a U at position 139, a G at position

173, an A at position 203, a U at position 226, a U at position 228, a C at position 270, a C at position 284, a C at position 299, a U at position 333, a G at position 337, a G at position 351, a C at position 362, a G at position 371, a G at position 386, any base at position 390, an A at position 396. Also preferred are RNase P RNA molecules that contain all of these alterations.

Also preferred are RNase P RNA molecules that contain one or more of the following alterations: a U at position 18, a U at position 59, a U at position 139, a G at position 173, a C at position 299, a G at position 371, an A at position 396. Also preferred are RNase P RNA molecules that contain all of these alterations. Other useful variant RNase P RNA molecules are those that contain one or more of the following alterations: a U at position 29, a U at position 136, a C at position 284, a U at position 333, a G at position 351. Also preferred are RNase P RNA molecules that contain all of these alterations.

Also preferred are RNase P RNA molecules that contain one or more of the following alterations: a U at position 18, a U at position 59, an A at position 137, a U at position 139, a G at position 173, an A at position 203, a U at position 226, a C at position 299, a G at position 337, a C at position 362, a G at position 371, an A at position 396. Also preferred are RNase P RNA molecules that contain all of these alterations.

Also preferred are RNase P RNA molecules that contain one or more of the following alterations: a C at position 18, an A at position 19, a U at position 26, a C at position 28, a U at position 59, an A at position 87, an A at position 101, a U at position 129, a U at position 138, a U at position 139, a G at position 173, a U at position 228, a C at position 270, a C at position 299, a G at position 371, a G at position 386, any base at position 390, an A at position 396. Also preferred are RNase P RNA molecules that contain all of these alterations.

Also preferred are RNase P RNA molecules that contain one or more of the following mutations: a C at position 18, a U at position 26, a U at position 59, a U at position 129, an U at position 138, a U at position 139, a

G at position 173, an A at position 203, a U at position 228, a C at position 270, a C at position 299, a G at position 371, and an A at position 396. Also preferred are RNase P RNA molecules that contain all of these alterations. Also preferred are RNase P RNA molecules that contain a U at one or both positions 136 and 333.

Every combination of any number of the above indicated alterations is specifically contemplated. Thus, RNase P RNA molecules containing two or more, three or more, four or more, five or more, six or more, etc., of the altered positions are specifically contemplated. Each of the preferred RNase P RNA molecules described herein as containing particular alterations can also contain additional alterations not listed. Those RNase P RNA molecules having only specific alterations relative to wild-type RNase P RNA are referred to herein as RNase P RNA molecules having only the specified alterations. Unless otherwise indicated, all references to nucleotide positions in RNase P RNA refer to the position numbering used in Figure 3. The positions referred to are those depicted in Figure 3. Deletions and insertions in any variant RNase P RNA does not change the reference position numbers of the original or remaining nucleotides.

Also preferred are RNase P RNA molecules containing alterations reducing substrate binding. Also preferred are RNase P molecules with an increased K_M. Also preferred are RNase P molecules with a destabilized guide sequence secondary structure. Also preferred are RNase P RNA molecules that maintain the P4 pseudoknot.

5. DNA Cleavage Efficiency

The disclosed variant RNase P RNA molecules cleave DNA substrates with greater efficiency than wild-type RNase P RNA. Methods for determining the cleavage activity of an RNase P RNA molecule are known, some of which are described herein. For example, one way is to compare the function of the variant RNase P RNA molecule to be tested to the activity of M1 RNA (wild-type RNase P RNA) as follows. First, the DNA oligonucleotide substrate is pre-annealed to an external (that is, untethered) RNA guide sequence identical to the guide sequence specific for

the substrate (Figure 4). Cleavage of this hybrid double-stranded helix is assayed in *trans* using native M1 RNA (that is, wild-type RNase P) and the variant RNase P RNAs of interest. There are numerous ways that these types of reactions can be performed. Note that no guide sequence would be used if the variant RNase P RNA molecule was selected to recognize the target molecule directly. An RNase P protein subunit can also be used when assessing cleavage.

Once the reactions are performed a kinetic analysis can be done to determine if the variant RNase P RNA cleaves DNA with a greater efficiency than wild-type. One way to determine efficiency is to compare the first order rate constants (for example $k_{cat,}$). Under single turnover conditions k_{cat} is a measure of the relative efficiency of the catalytic step because product release is often rate limiting in an RNase P catalytic reaction (Tallsjo & Kirsebom, *Nucleic Acids Res* 21(1):51-57 (1993); Herschlag, *Proc Natl Acad Sci U S A* 88(16):6921-6925 (1991); and Herschlag & Cech, *Biochemistry* 29(44):10159-10171 (1990)). Techniques for performing an RNase P reaction under single turnover conditions, by for example having a vast excess of GS bound substrate relative to RNase P RNA, are known and can be used with the disclosed variant RNase P RNA molecules. The presence, absence, or concentration of RNase P protein subunit can also be used to assess the kinetics of cleavage by the variant RNase P RNA molecules.

It is preferred that the variant RNase P RNA have a k_{cat} at least 10 times greater than the wild-type RNase P RNA. It is more preferred that the variant RNase P RNA have a k_{cat} at least 100 times greater than the wild-type RNase P RNA. It is more preferred that the variant RNase P RNA have a k_{cat} at least 1000 times greater than the wild-type RNase P RNA. It is more preferred that the variant RNase P RNA have a k_{cat} at least approximately 250 times greater than the wild-type RNase P RNA. It is most preferred that the variant RNase P RNA have a k_{cat} at least approximately 7000 times greater than the wild-type RNase P RNA. An example of the comparison of

variant and wild-type RNase P RNA DNA-cleaving activity is described in Example 3.

Another way to kinetically analyze variant RNase P RNA and wild-type RNase P RNA is to compare the k_{cat}/K_M. It is preferred that the k_{cat}/K_M of variant RNase P RNA is at least 10 fold greater than the k_{cat}/K_M of the wild-type RNase P RNA. It is more preferred that the k_{cat}/K_M of variant RNase P RNA is at least 100 fold greater than the k_{cat}/K_M of the wild-type RNase P RNA. It is more preferred that the k_{cat}/K_M of variant RNase P RNA is at least 1000 fold greater than the k_{cat}/K_M of the wild-type RNase P RNA. It is more preferred that the k_{cat}/K_M of variant RNase P RNA is at least approximately 136 fold greater than the k_{cat}/K_M of the wild-type RNase P RNA. It is most preferred that the k_{cat}/K_M of variant RNase P RNA is at least approximately 1840 fold greater than the k_{cat}/K_M of the wild-type RNase P RNA.

Another way to kinetically analyze variant RNase P RNA and wild-type RNase P RNA is to compare the k_{obs}. It is preferred that the k_{obs} of variant RNase P RNA is at least 10 fold greater than the k_{obs} of the wild-type RNase P RNA. It is more preferred that the k_{obs} of variant RNase P RNA is at least 100 fold greater than the k_{obs} of the wild-type RNase P RNA. It is more preferred that the k_{obs} of variant RNase P RNA is at least 1000 fold greater than the k_{obs} of the wild-type RNase P RNA.

Another way to analyze the kinetic efficiency of variant RNase P RNA and wild-type RNase P RNA is to determine the specificity of the RNAs for a DNA substrate versus an RNA substrate. For example, a k_{cat}/K_M is determined for both an RNA substrate and a DNA substrate. The ratio of the k_{cat}/K_M for the DNA substrate to the k_{cat}/K_M of the RNA substrate is obtained. The higher this ratio is the more the RNase P prefers DNA as a substrate to RNA. After obtaining the DNA:RNA substrate ratio for both a variant RNase P RNA and wild-type RNase P RNA these values can be compared. It is preferred that variant RNase P RNA has a DNA:RNA ratio at least 10 times greater than the DNA:RNA ratio of wild-type RNase P RNA. It is more preferred that variant RNase P RNA has a DNA:RNA ratio

at least 100 times greater than the DNA:RNA ratio of wild-type RNase P RNA. It is more preferred that variant RNase P RNA has a DNA:RNA ratio at least 1000 times greater than the DNA:RNA ratio of wild-type RNase P RNA. It is more preferred that variant RNase P RNA has a DNA:RNA ratio at least approximately 800 times greater than the DNA:RNA ratio of wild-type RNase P RNA performed on an RNA oligonucleotide. It is most preferred that variant RNase P RNA has a DNA:RNA ratio at least approximately 2400 times greater than the DNA:RNA ratio of wild-type RNase P RNA. It is understood that this type of kinetic comparison can be performed with a number of different RNA substrates. For example an RNA oligonucleotide or p4.5S RNA or ptRNA can be used as an RNA substrate.

It is preferred that the variant RNase P RNA molecules have an increased catalytic efficiency over wild-type RNase P RNA when the $MgCl_2$ concentrations are decreased. It is more preferred that the variant RNase P RNA molecules have an increased catalytic efficiency over wild-type RNase P RNA when the $MgCl_2$ concentrations are less than 100 mM. It is more preferred that the variant RNase P RNA molecules have an increased catalytic efficiency over wild-type RNase P RNA when the $MgCl_2$ concentrations are less than 50 mM. It is most preferred that the variant RNase P RNA molecules have an increased catalytic efficiency over wild-type RNase P RNA when the $MgCl_2$ concentrations are less than 20 mM. The variant RNase P RNA molecules can also have an increased catalytic efficiency over wild-type RNase P RNA in the presence of RNase P protein subunit.

B. Guide Sequences

RNase P recognizes structures in substrate molecules for cleavage to take place. Natural substrates of RNase P, precursor tRNA molecules and p4.5S RNA, have stem and loop structures that are recognized by RNase P as substrate structures. Cleavage of other substrates by RNase P has been achieved through the use of guide sequences. Guide sequences are oligonucleotides that, in combination with a target sequence, form a structure recognized by RNase P as a substrate. The requirements for substrates of

eukaryotic and prokaryotic RNase P differ and so the requirements of guide sequences for use with eukaryotic and prokaryotic RNase P differ.

Since RNase P RNA recognition of substrates is primarily recognition of structure, variant RNase P RNA molecules can also be selected to recognize and cleave target molecules directly without the use of a guide sequence. A preferred example is variant RNase P RNA that directly recognizes and cleaves double stranded DNA. Other nucleic acid structures can also be used as direct targets for variant RNase P RNA molecules.

Guide sequences for use in the method to select variant RNase P RNA molecules are preferably designed to cleave an RNA substrate the corresponds to the DNA substrate of interest. As used herein, an RNA substrate, molecule, or sequence is said to correspond to a DNA substrate, molecule, or sequence when the sequence of the RNA corresponds to the sequence of the DNA. As used herein, correspondence of the sequence of an RNA and a DNA means that the sequences are the same except that uridines in the RNA are thymidines in the DNA, and vice versa. Thymidine and uridine are "corresponding" nucleotides.

1. Prokaryotic Guide Sequences

The requirements for a guide sequence functional with wild-type prokaryotic RNase P are (1) nucleotide sequence which specifically binds to the targeted substrate to produce a short sequence of base pairs 3' to the cleavage site on the substrate and (2) a terminal 3'-NCCA, where N is any nucleotide, preferably a purine. This is a minimum structure and can include additional elements. Such additional elements can include structures similar to any or all of the stem and loop structures of a tRNA molecule. In particular, prokaryotic RNase P can use any eukaryotic guide sequence (the structures of which are described below). For the disclosed variant RNase P RNA molecules, guide sequences forming only the stem structure can be used. The selection process can eliminate the requirement for a terminal 3'-NCCA.

Preferred prokaryotic guide sequences for use in the disclosed methods generally have no fewer than four, and more usually six to fifteen,

nucleotides complementary to the targeted sequence. It is not critical that all nucleotides be complementary, although the efficiency of the reaction will vary with the degree of complementarity. Alternatively, the variant RNase P RNA can be selected to efficiently cleave incompletely base paired target sequences. The rate of cleavage is dependent on the RNase P, the secondary structure of the hybrid substrate, which includes the targeted sequence and the presence of the 3'-NCCA in the hybrid substrate. The design and structure of prokaryotic guide sequences for use with wild-type RNase P RNA is described in U.S. Patent No. 5,168,053 to Altman et al. Such guide sequences can be used or adapted for use with the disclosed variant RNase P RNA molecules.

2. Eukaryotic Guide Sequences

A guide sequence for wild-type eukaryotic RNase P consists of a sequence which, when in a complex with the target sequence, forms a secondary structure resembling that of a tRNA cloverleaf or a part of it. As used herein, the term "resembling a precursor tRNA" means a complex formed by the GS with target sequence to resemble a sufficient portion of the tRNA structure to result in cleavage of the target DNA by RNase P. The guide sequence can be derived from any tRNA except that the D stem and aminoacyl stem should be altered to be complementary to the target substrate sequence. These altered stems are referred to as recognition arms. The recognition arm corresponding to the aminoacyl stem is referred to as the A recognition arm and the recognition arm corresponding to the D stem is referred to as the D recognition arm.

Between the A recognition arm and the D recognition arm on a eukaryotic guide sequence is a sequence that can form a stem loop structure. The presence of a 3'-CCA on a guide sequence enhances the efficiency of *in vitro* reaction with the eukaryotic RNase P but is not required. Eukaryotic guide sequences can, but need not, have sequences that can form structures similar to the anticodon loop and stem and extra loop of tRNA. Elimination of the anticodon stem and loop increases the efficiency of the reaction by

about ten fold. Changes in other parts of an GS can increase its efficiency about one hundred fold.

The desired secondary structure is determined using conventional Watson-Crick base pairing schemes to form a structure resembling a tRNA, that is, having structure as described below. The specific sequence of the hydrogen bonded regions is not as critical, as long as the desired structure is formed. All tRNAs, including tRNAs from a wide variety of bacteria and eukaryotes, conform to the same general secondary structure. This is typically written in the form of a cloverleaf, maintained by hydrogen-bonded base pairing between short complementary regions. The four major arms are named for their structure or function: The acceptor arm consists of a 3' terminal CCA_{OH} plus a variable fourth nucleotide extending beyond the stem formed by base-pairing the 5' and 3' segments of the molecule. The other arms consist of base-paired stems and unpaired loops. The "T" arm is named for the presence of the ribothymidine nucleotide and contains seven unpaired bases in the loop. The anticodon arm always contains the anticodon triplet in the center of the loop and consists of seven unpaired bases. The D arm is named for the presence of the base dihydrouridine in the loop, another of the chemically modified bases in tRNA, and includes between eight and twelve unpaired bases. Positions are numbered from 5' to 3' according to the most common tRNA structure, which has 76 residues. The overall range of tRNA lengths is from 74 to 95 bases. The variation in length is caused by differences in the structure of two of the arms, the D arm and the extra or variable arm, which lies between the T and anticodon arms, which can contain between three and five bases, or between 13 and 21 bases with a stem of about five bases. The base pairing that maintains the secondary structure is virtually invariant: there are always seven base pairs in the acceptor stem, five in the T arm, five in the anticodon arm, and three or four in the D arm.

Guide sequences for use with RNase P RNA, including the disclosed variant RNase P RNA, preferably include a D stem, an aminoacyl stem, and a T loop and stem. For the disclosed variant RNase P RNA molecules, guide

sequences forming only the stem structure can also be used. The selection process can eliminate the requirement for additional structure when the variant RNase P RNA is based on eukaryotic RNase P.

The complementary sequences will preferably consist of eleven nucleotides, or, under certain conditions may consist of as few as seven nucleotides, in two blocks which base pair with the target sequence and which are separated by two unpaired nucleotides in the target sequence, preferably UU, wherein the two blocks are complementary to a sequence 3' to the site targeted for cleavage. The design and structure of eukaryotic guide sequences for use with wild-type RNase P are described in U.S. Patent Nos. 5,624,824 and 5,869,248 to Yuan et al. Such guide sequences can be used or adapted for use with the disclosed variant RNase P RNA molecules.

3. Guide Sequence Modifications

The guide sequence can be made of ribonucleotides, deoxyribonucleotides, or a combination. The guide sequence can also contain chemically modified nucleotides. Many modifications are known (see, for example, European Patent Application No. 96902682.2). Preferred chemical modifications are those that promote strand invasion by the guide sequence. It is preferred that the guide sequence be composed of ribonucleotides. Use of a deoxyribonucleotide guide sequences for selection of variant RNase P RNA molecules can result in variant RNase P RNA that can recognize and cleave a double stranded DNA molecule without the need for a separate guide sequence.

A guide sequence and RNase P RNA can be linked to form a single oligonucleotide molecule possessing both the targeting function of an GS and cleavage function of RNase P RNA. Such a combination, in a single oligonucleotide molecule, is referred to as an RNase P internal guide sequence (RIGS). An RIGS can be used to cleave a target DNA molecule in the same manner as separate GS and RNase P RNA. Use of a RIGS for cleavage is preferred.

RIGSs can be formed by linking a guide sequence to an RNase P RNA by any suitable means. For example, a GS and RNase P RNA can be

prepared as separate molecules which are then covalently linked *in vitro*. Alternatively, a complete RIGS can be synthesized as a single molecule, either by chemical synthesis, or by *in vitro* or *in vivo* transcription of a DNA molecule encoding linked GS and RNase P RNA. The linkage between the GS and RNase P domains of an RIGS can have any form that allows the domains to cleave a target DNA. For example, the two domains could be joined by an oligonucleotide linker or tether. Preferably, the linker will be composed of an ordinary nucleotides joined by phosphodiester bonds. Use of an oligonucleotide tether is illustrated in the examples. The GS and RNase P RNA components can be joined in either order, with the RNase P RNA linked to either the 3' end or 5' end of the GS component.

RIGSs can be used for cleavage of target DNA both *in vitro* and *in vivo*. *In vitro*, the RIGS can function without RNase P protein components, although activity of the RIGS can be increased by the addition of RNase P protein components. *In vivo*, endogenous RNase P proteins can stimulate activity of the RIGS. The activity of both prokaryotic- and eukaryotic-based RIGSs are expected to be enhanced by the presence of either prokaryotic or eukaryotic RNase P protein components.

C. Substrates and Targets

The disclosed variant RNase P RNA molecules are selected and used for cleavage of DNA molecules. Any DNA sequence of interest can be used as a target for cleavage by the disclosed variant RNase P RNA molecules. In general, any variant RNase P RNA can be targeted to the DNA sequence of interest by using the RNase P in association with a guide sequence that will interact with the DNA sequence. The requirements for design of guide sequences that interact with sequences of interest are discussed above. The disclosed variant RNase P RNA molecules can also be used without a guide sequence if the variant was selected to recognize a target molecule directly.

For selection of variant RNase P RNA molecules, the following features of substrate DNA are pertinent. It is preferred that the substrate be a single stranded oligodeoxyribonucleotide. In general, the substrate DNA should be designed to hybridize a guide sequence such that a structure

recognized by RNase P RNA is formed. The requirements of such structures are discussed above. Alternatively, substrate DNA can be directly targeted if the variant RNase P RNA was selected to recognize the substrate DNA directly. The substrate DNA should include a region that will hybridize to the guide sequence and a region to be cleaved off. The region to be cleaved off must be at the 5' end of the substrate DNA. The substrate DNA can also be associated with a capture tag to facilitate selection of variant RNase P RNA molecules as described below. It is preferred that the capture tag be coupled to the region to be cleaved off (that is, the 5' region of the substrate DNA).

The substrate DNA used in the selection method can be a separate molecule, which is preferred, or it can be covalently coupled to the guide sequence or RNase P. Different configurations are useful in different selection schemes. Preferably, the substrate DNA used in the selection method is modeled on a target DNA sequence of interest.

D. Capture Tags

Capture tags can be used to separate molecules which have a capture tag away from molecules which do not. In the context of the disclosed method, capture tags can be used to separate the two halves of a cleaved substrate DNA. In this way, variant RNase P RNA molecules that successfully cleave the substrate DNA can be separated from RNase P RNA molecules that do not.

As used herein, a capture tag is any compound that can be associated with a synthesized DNA molecule and which can be used to separate compounds or complexes having the capture tag from those that do not. Preferably, a capture tag is a compound, such as a ligand or hapten, that binds to or interacts with another compound, such as a ligand-binding molecule or an antibody. The other compound with which the capture tag can associate is referred to as a capture tag receptor (CTR). It is also preferred that such interaction between the capture tag and the capturing component be a specific interaction, such as between a hapten and an antibody or a ligand and a ligand-binding molecule.

Suitable capture tags include hapten or ligand molecules that can be coupled to the 5' end of the synthesized DNA molecule. Preferred capture tags, described in the context of nucleic acid probes, have been described by Syvanen *et al.*, *Nucleic Acids Res.*, 14:5037 (1986)). Preferred capture tags include biotin, which can be incorporated into nucleic acids (Langer *et al.*, *Proc. Natl. Acad. Sci. USA* 78:6633 (1981)) and captured using streptavidin or biotin-specific antibodies. A preferred hapten for use as a capture tag is digoxygenin (Kerkhof, *Anal. Biochem.* 205:359-364 (1992)). Many compounds for which a specific antibody is known or for which a specific antibody can be generated can be used as capture tags. Such capture tags can be captured by antibodies which recognize the compound. Antibodies useful as capture tags can be obtained commercially or produced using well established methods. For example, Johnstone and Thorpe, *Immunochemistry In Practice* (Blackwell Scientific Publications, Oxford, England, 1987), on pages 30-85, describe general methods useful for producing both polyclonal and monoclonal antibodies.

Another preferred capture tag is an anti-antibody antibody. Such anti-antibody antibodies and their use are well known. For example, anti-antibody antibodies that are specific for antibodies of a certain class (for example, IgG, IgM), or antibodies of a certain species (for example, anti-rabbit antibodies) are commonly used to detect or bind other groups of antibodies. Thus, one can have an antibody to the capture tag and then this antibody:capture tag:DNA complex can then be captured by binding to an antibody for the antibody portion of the complex.

Another form of capture tag is one which can form selectable cleavable covalent bonds with other molecules of choice. For example, a preferred capture tag of this type is one which contains a sulfur atom. A DNA molecule which is associated with this capture tag can be captured by retention on a thiolpropyl sepharose column. Extensive washing of the column removes unwanted molecules and reduction with β-mercaptoethanol, for example, allows the desired DNA molecules to be collected after purification under relatively gentle conditions (See Lorsch and Szostak, 1994).

E. Supports

Capture tags can be associated with the substrate DNA and then the capture tag:DNA complex can be selectively captured through interaction with a capture tag receptor (CTR). CTRs can be associated with a solid support. When capture tag:DNA complexes are bound to CTRs of this type they can be effectively separated from the other portion of the substrate DNA when cleavage occurs. In this way, variant RNase P RNA molecules that successfully cleave the substrate DNA can be separated from RNase P RNA molecules that do not.

CTRs can be coupled to can be any solid material to which the CTRs can be adhered or coupled. This includes materials such as acrylamide, cellulose, nitrocellulose, glass, polystyrene, polyethylene vinyl acetate, polypropylene, polymethacrylate, polyethylene, polyethylene oxide, glass, polysilicates, polycarbonates, teflon, fluorocarbons, nylon, silicon rubber, polyanhydrides, polyglycolic acid, polylactic acid, polyorthoesters, polypropylfumerate, collagen, glycosaminoglycans, and polyamino acids. Supports can have any useful form including thin films or membranes, beads, bottles, dishes, fibers, woven fibers, shaped polymers, particles and microparticles. Preferred forms of supports are plates and beads. The most preferred form of beads are magnetic beads.

Methods for immobilizing antibodies to solid-state substrates are well established. Immobilization can be accomplished by attachment, for example, to aminated surfaces, carboxylated surfaces or hydroxylated surfaces using standard immobilization chemistries. Examples of attachment agents are cyanogen bromide, succinimide, aldehydes, tosyl chloride, avidin-biotin, photocrosslinkable agents, epoxides and maleimides. A preferred attachment agent is glutaraldehyde. These and other attachment agents, as well as methods for their use in attachment, are described in *Protein immobilization: fundamentals and applications*, Richard F. Taylor, ed. (M. Dekker, New York, 1991), Johnstone and Thorpe, *Immunochemistry In Practice* (Blackwell Scientific Publications, Oxford, England, 1987) pages 209-216 and 241-242, and *Immobilized Affinity Ligands*, Craig T. Hermanson *et al.*, eds. (Academic

Press, New York, 1992). Antibodies can be attached to a support by chemically cross-linking a free amino group on the antibody to reactive side groups present within the solid-state support. For example, antibodies may be chemically cross-linked to a support that contains free amino or carboxyl groups using glutaraldehyde or carbodiimides as cross-linker agents. In this method, aqueous solutions containing free antibodies are incubated with the solid-state substrate in the presence of glutaraldehyde or carbodiimide. For crosslinking with glutaraldehyde the reactants can be incubated with 2% glutaraldehyde by volume in a buffered solution such as 0.1 M sodium cacodylate at pH 7.4. Other standard immobilization chemistries are known by those of skill in the art.

In addition, non-antibody proteins such as streptavidin, can be linked using similar methods. Many protein and antibody columns are commercially available as well as specifically derivatized supports for conjugation to the CTRs.

Methods for immobilization of oligonucleotides to substrates are well established. Oligonucleotides, including oligonucleotide capture tag receptors, can be coupled to substrates using established coupling methods. For example, suitable attachment methods are described by Pease *et al.*, *Proc. Natl. Acad. Sci. USA* 91(11):5022-5026 (1994), and Khrapko *et al.*, *Mol Biol (Mosk) (USSR)* 25:718-730 (1991). A method for immobilization of 3'-amine oligonucleotides on casein-coated slides is described by Stimpson *et al.*, *Proc. Natl. Acad. Sci. USA* 92:6379-6383 (1995). A preferred method of attaching oligonucleotides to solid-state substrates is described by Guo *et al.*, *Nucleic Acids Res.* 22:5456-5465 (1994).

II. Methods

The disclosed variant RNase P RNA molecules are made by a process of *in vitro* selection. Any *in vitro* selection system can be used for this purpose. Many such systems are known, some of which are described below. The resulting variant RNase P RNA molecules can be used to cleave any DNA sequence of interest in a variety of contexts. Such cleavage can be used in a variety of methods, including diagnostic methods for detection,

quantitation, or cataloging of DNA sequences, forensic methods, genome dissection methods, biostatistical methods, and population genetics methods, cleaving genomic DNA at particular sequences, creating gene knockouts by gene cleavage, killing specific cells by specific cleavage of DNA, cleaving pathogen DNA in a host cell, and killing mutant cells by specific cleavage of mutant DNA in the cell.

A. Selection of Variant RNase P RNA Molecules

Preferred forms of the method for selection of variant RNase P RNA molecules involve the following steps: (1) generating a population of variant RNase P RNAs, where the sequence of the variant RNase P RNAs differ from the sequence of wild-type RNase P RNA, (2) mixing the population of variant RNase P RNAs with a DNA substrate, and (3) separating variant RNAse P RNAs that cleave the DNA substrate from variant RNAse P RNAs that do not. In general, steps (2) and (3) should be repeated multiple times using the population of variant RNase P RNAs that cleave to generate a new population. Also in general, after several repetitions of steps (2) and (3), additional variations can be introduced into the selected population of variant RNase P RNAs that cleave. If desired, the method, and in particular step (2), can be carried out in the presence of RNase P protein subunit.

1. *In vitro* Selection

Oligonucleotide molecules with a given function, catalytic or ligand-binding, can be isolated from a complex mixture of random oligonucleotides in what has been referred to as "in vitro genetics" (Szostak, *TIBS* 19:89, 1992). One synthesizes a large set of molecules bearing random and defined sequences and subjects that complex mixture, for example, approximately 10^{15} individual sequences in 100 μg of a 100 nucleotide RNA, to some selection and enrichment process. Through repeated cycles of affinity chromatography and PCR amplification of the molecules bound to the ligand on the column, Ellington & Szostak, *Nature* 346(6287), 818-822 (1990), estimated that 1 in 10^{10} RNA molecules folded in such a way as to bind a small molecule dyes. DNA molecules with such ligand-binding behavior have been isolated as well (Ellington and Szostak, 1992; Bock et al, 1992).

Techniques aimed at similar goals exist for small organic molecules, proteins, antibodies and other macromolecules known to those of skill in the art. Screening sets of molecules for a desired activity whether based on small organic libraries, oligonucleotides, or antibodies is broadly referred to as combinatorial chemistry.

2. Selection of Variant RNase P RNAs Having Enhanced DNA Cleaving Activity

There are three general *in vitro* selection methods which can be used to isolate variant RNase P RNAs which cleave DNA with an increased catalytic efficiency: (1) cis, (2) cis-trans, and (3) trans.

These three general methods have a number of elements in common and certain elements which are unique to each. It is preferred that each method have all of these elements, but not all of these elements are required for the methods to function. For example, all methods require a period whereby the variant RNase P RNA molecules (the pool) react with the substrate. All methods require a way to amplify the RNase P RNA molecules. All methods require a way to separate the molecules with a desired activity (i.e. DNA cleavage) from those molecules that do not function as desired. All methods are preferably performed iteratively. All methods require a "pool" of RNase P RNA molecules. This set of RNase P RNA molecules is nothing more than at least two (but preferably many more) RNase P RNA molecules that have an altered sequence relative to wild-type RNase P RNA at least at one position.

The conditions under which the reactions can be performed may be varied and this allows for the stringency of the method to be controlled. For example, the Mg^{++} concentration can be altered from round to round during the *in vitro* selection. All methods generally require the set of RNase P RNA molecules to go through a reaction step, a selection step which separates the functional molecules, and then an amplification step of the new set of RNase P RNA molecules which are enriched with molecules that function as desired, and lastly the performance of another cycle of the above three steps.

Each of these methods is discussed in detail below, focusing on the differences between each type.

The set of variant RNase P RNA molecules can be obtained in many different ways. All that is required is that sequence variation be introduced into a population of RNase P RNA molecules. For example, through the use of solid state chemical DNA synthesis, each position of the wild type RNase P can be specifically varied. There are a number of different solid state chemical DNA synthesis methods, all of which can be used to create the sets of RNase P RNAs or the molecules that encode them, but the phosphoramidite method is preferred.

Very long oligonucleotides can be synthesized using this method (greater than 200 bases) and sets of DNA molecules with greater than 10^{17} members can be obtained depending on the length of the oligonucleotide. Furthermore, by appropriately mixing the phosphoramidite reagents, molecules which are completely randomized at a given position can be generated as well as molecules that vary by a specific amount for example 3 % mutagenesis. This type of DNA synthesis is described by Green et al., *Methods: A Companion to Methods in Enzymology* 2:75-86 (1991).

Since it would be difficult to chemically synthesize a DNA molecule as long as the RNase P RNA, a preferred method would be to synthesize for example, three cassettes, representing the complete RNase P RNA molecule and then through recombinant molecular biology techniques and PCR generate a complete RNase P RNA encoding set. Methods such as these can be found for example, in Bartel & Szostak, *Science* 261:1411-1416 (1993).

The set of molecules can also be generated by performing mutagenic PCR. Heterogeneous populations of molecules can be generated using for example the methods of Cadwell, & Joyce, *PCR Methods Appl* 3(6):136-140 (1994). In general, PCR is performed in conditions which decrease the fidelity of the thermal stable polymerase during chain extension. Common techniques for doing this are, for example, decreasing or increasing the Mg^{++} concentration of the reaction or the monovalent salt concentration of the

reaction. This type of method produces a set of molecules that usually have up to 7% mutagenesis across the sequence that is amplified.

It is understood that when performing these methods that one can easily move in either direction between the RNase P RNA and the DNA that encodes the RNase P RNA by using PCR, *in vitro* transcription reactions, and *in vitro* reverse transcription reactions.

In general amplification of the compositions and sets of molecules occurs through PCR and transcription reactions are performed to obtain the RNase P RNA molecules, whether wild-type or variant. It is understood that in general the sets of molecules that encode RNase P RNA molecules should contain defined sequence for PCR amplification and appropriate transcription start sites, such as a T7 polymerase start site.

a. Cis system

The base *cis* system consists of a set of modified RNase P RNA molecules, a guide sequence (if used) covalently coupled to the 3'-end of the RNA molecules that make up the set, and a DNA substrate covalently to the 3'-end of the guide sequence (or to the RNase P RNA). In addition, a means for isolating the molecules that function as desired is required. Suitable means for isolating in the *cis* system would be for example, gel chromatography, affinity column chromatography, and size exclusion chromatography. The key aspect of the *cis* system is that all three components, the RNase P RNA, the guide sequence (if used), and the substrate are covalently attached together. RNase P RNAs that cleave the substrate will in essence cleave themselves producing two products: the RNase P RNA attached to the guide sequence and part of the substrate, and a free oligonucleotide which will correspond to the portion of the substrate which is on the 3' side of the cleavage site. If RNase P protein subunit is used, it can be, but preferably is not, covalently attached to the other components.

This type of system is amenable to numerous techniques of selectively retrieving the RNase P RNA molecules which have increased efficiency over wild-type RNase P RNA for cleaving DNA substrates. For

example, the 3' end of the RNase P:GS:Substrate molecule can be attached to a biotin molecule. After cleavage, those molecules that function as desired (i.e. enhanced cleavage of the substrate) will lose the biotin capture tag and will not be selectively retained on a streptavidin or avidin affinity column. As discussed above this type of selection could be performed with numerous different capture tags. The functional molecules could also be separated by polyacrylamide or agarose gel chromatography. Functional molecules that have cleaved off a portion of the substrate will be shorter than non-functional molecules, and can be separated and recovered on this basis.

b. Cis-Trans system

The base *cis-trans* system consists of a set of modified RNase P RNA molecules, a guide sequence covalently coupled to the 3'-end of the RNA molecules that make up the set, and a free DNA substrate which is not attached to the guide sequence (but which, of course, can hybridize to the guide sequence). In addition, a means for isolating the molecules that function as desired is required. Suitable means for isolating for the *cis-trans* system would be for example, gel chromatography, affinity column chromatography, and size exclusion chromatography. A *cis-trans* system of selection is described in the examples.

c. Trans system

The base *trans* system consists of a set of modified RNase P RNA molecules, a free guide sequence (if used) not attached to the RNase P RNA, and a free DNA substrate which is not attached to the guide sequence (or to the RNase P RNA). In addition, a means for isolating the molecules that function as desired is required. Suitable means for isolating for the *cis-trans* system would be for example, gel chromatography, affinity column chromatography, and size exclusion chromatography.

3. Methods of Using Variant RNase P RNA Molecules

The disclosed variant RNase P RNA molecules have a wide variety of uses. For example, the disclosed variant RNase P RNA molecules can be used to cleave any desired DNA sequence *in vitro* or *in vivo*. These uses specifically include diagnostic methods for detection, quantitation, or

cataloging of DNA sequences, forensic methods, genome dissection methods, biostatistical methods, and population genetics methods, cleaving genomic DNA at particular sequences, creating gene knockouts by gene cleavage, killing specific cells by specific cleavage of DNA, cleaving pathogen DNA in a host cell, and killing mutant cells by specific cleavage of mutant DNA in the cell. The disclosed variant RNase P RNA molecules can be used in the presence or absence of RNase P protein subunits.

Use of the disclosed variant RNase P RNA molecules for preventing or altering expression of a gene *in vivo* has advantages over most other approaches to such regulation (for example, antisense RNA and transcription inhibitors) since the disclosed molecules target and inactivate the gene. In this way, the cell cannot circumvent regulation by increasing the number of transcripts of the gene as is the potential with other forms of regulation. The disclosed molecules have the additional advantage in that they can attack dormant (that is, non-expressed) genes while many other forms of regulation require gene expression to be effective. Thus, the disclosed molecules can be particularly effective in attacking dormant viruses in cells.

A major area of use for the disclosed RNase P RNA molecules is for in vitro cleavage of DNA. Many cloning, diagnostic, genomic, and biostatistical methods rely on manipulation of DNA. Where specific cleavage of DNA is useful in a method, the disclosed variant RNase P RNA molecules can be used. In general, all that is required is the association with the RNase P RNA of an appropriate guide sequence targeted to the sequence to be cleaved. Alternatively, a variant RNase P RNA selected to recognize and cleave a particular sequence can be used. In either case the RNase P RNA serves as a restriction enzyme. The disclosed variant RNase P RNA molecules are more useful than restriction enzymes, however, since they can be targeted to any sequence. The disclosed variant RNase P RNA molecules are also more useful, especially for manipulations of genomic DNA, since the recognition sequence (that is, the length of the guide sequence/target hybrid) is longer than the recognition sequence of most restriction enzymes.

4. Delivery of Variant RNase P RNA Molecules to Cells

The disclosed variant RNase P RNA molecules can be delivered to and/or introduced into cells by a variety of methods. For example, the variant RNase P RNA can be introduced into cells by direct transfer of genetic material, in a plasmid or viral vector, or via transfer of genetic material in cells or carriers such as cationic liposomes. Such methods are well known in the art and can be readily adapted for use with the disclosed variant RNase P RNA molecules. Further, these methods can be used to target certain diseases and cell populations by using the targeting characteristics of the carrier. Transfer vectors can be any nucleic acid construct used to deliver genes into cells (for example, a plasmid), or as part of a general strategy to deliver nucleic acids or genes, for example, as part of recombinant retrovirus or adenovirus (Ram et al., *Cancer Res.* 53:83-88, (1993)). Appropriate means for transfection, including viral vectors, chemical transfectants, or physico-mechanical methods such as electroporation and direct diffusion of DNA, are described by, for example, Wolff, J. A., et al., *Science*, 247:1465-1468, (1990); and Wolff, J. A., *Nature*, 352:815-818, (1991).

a. Vectors

As used herein, plasmid or viral vectors are agents that transport a nucleic acid into the cell without degradation and include a promoter yielding expression of the gene in the cells into which it is delivered. In a preferred embodiment vectors are derived from either a virus or a retrovirus. Preferred viral vectors are Adenovirus, Adeno-associated virus, Herpes virus, Vaccinia virus, Polio virus, AIDS virus, neuronal trophic virus, Sindbis and other RNA viruses, including these viruses with the HIV backbone. Also preferred are any viral families which share the properties of these viruses which make them suitable for use as vectors. Preferred retroviruses include Murine Maloney Leukemia virus, MMLV, and retroviruses that express the desirable properties of MMLV as a vector. Retroviral vectors are able to carry a larger genetic payload, i.e., a transgene or marker gene, than other viral vectors, and for this reason are a commonly used vector. However,

they are not useful in non-proliferating cells. Adenovirus vectors are relatively stable and easy to work with, have high titers, and can be delivered in aerosol formulation, and can transfect non-dividing cells. Pox viral vectors are large and have several sites for inserting genes, they are thermostable and can be stored at room temperature. A preferred embodiment is a viral vector which has been engineered so as to suppress the immune response of the host organism, elicited by the viral antigens. Preferred vectors of this type will carry coding regions for Interleukin 8 or 10.

Viral vectors have higher transaction abilities (ability to introduce genes) than do most chemical or physical methods to introduce genes into cells. Typically, viral vectors contain, nonstructural early genes, structural late genes, an RNA polymerase III transcript, inverted terminal repeats necessary for replication and encapsidation, and promoters to control the transcription and replication of the viral genome. When engineered as vectors, viruses typically have one or more of the early genes removed and a gene or gene/promotor cassette is inserted into the viral genome in place of the removed viral DNA. Constructs of this type can carry up to about 8 kb of foreign genetic material. The necessary functions of the removed early genes are typically supplied by cell lines which have been engineered to express the gene products of the early genes in trans.

Retroviral Vectors: A retrovirus is an animal virus belonging to the virus family of Retroviridae, including any types, subfamilies, genus, or tropisms. Retroviral vectors, in general, are described by Verma, I.M., Retroviral vectors for gene transfer. In MICROBIOLOGY-1985, American Society for Microbiology, pp. 229-232, Washington, (1985). Examples of methods for using retroviral vectors for gene therapy are described in U.S. Patent Nos. 4,868,116 and 4,980,286; PCT applications WO 90/02806 and WO 89/07136; and Mulligan, (*Science* 260:926-932 (1993)).

A retrovirus is essentially a package which has packed into it nucleic acid cargo. The nucleic acid cargo carries with it a packaging signal, which ensures that the replicated daughter molecules will be efficiently packaged

within the package coat. In addition to the package signal, there are a number of molecules which are needed in *cis*, for the replication, and packaging of the replicated virus. Typically a retroviral genome, contains the gag, pol, and env genes which are involved in the making of the protein coat. It is the gag, pol, and env genes which are typically replaced by the foreign DNA that it is to be transferred to the target cell. Retrovirus vectors typically contain a packaging signal for incorporation into the package coat, a sequence which signals the start of the gag transcription unit, elements necessary for reverse transcription, including a primer binding site to bind the tRNA primer of reverse transcription, terminal repeat sequences that guide the switch of RNA strands during DNA synthesis, a purine rich sequence 5' to the 3' LTR that serve as the priming site for the synthesis of the second strand of DNA synthesis, and specific sequences near the ends of the LTRs that enable the insertion of the DNA state of the retrovirus to insert into the host genome. The removal of the gag, pol, and env genes allows for about 8 kb of foreign sequence to be inserted into the viral genome, become reverse transcribed , and upon replication be packaged into a new retroviral particle. This amount of nucleic acid is sufficient for the delivery of a one to many genes depending on the size of each transcript. It is preferable to include either positive or negative selectable markers along with other genes in the insert.

Since the replication machinery and packaging proteins in most retroviral vectors have been removed (gag, pol, and env), the vectors are typically generated by placing them into a packaging cell line. A packaging cell line is a cell line which has been transfected or transformed with a retrovirus that contains the replication and packaging machinery, but lacks any packaging signal. When the vector carrying the DNA of choice is transfected into these cell lines, the vector containing the gene of interest is replicated and packaged into new retroviral particles, by the machinery provided in *cis* by the helper cell. The genomes for the machinery are not packaged because they lack the necessary signals.

Adenoviral Vectors: The construction of replication-defective adenoviruses has been described (Berkner et al., *J. Virology* 61:1213-1220 (1987); Massie et al., *Mol. Cell. Biol.* 6:2872-2883 (1986); Haj-Ahmad et al., *J. Virology* 57:267-274 (1986); Davidson et al., *J. Virology* 61:1226-1239 (1987); Zhang "Generation and identification of recombinant adenovirus by liposome-mediated transfection and PCR analysis" *BioTechniques* 15:868-872 (1993)). The benefit of the use of these viruses as vectors is that they are limited in the extent to which they can spread to other cell types, since they can replicate within an initial infected cell, but are unable to form new infectious viral particles. Recombinant adenoviruses have been shown to achieve high efficiency gene transfer after direct, *in vivo* delivery to airway epithelium, hepatocytes, vascular endothelium, CNS parenchyma and a number of other tissue sites (Morsy, *J. Clin. Invest.* 92:1580-1586 (1993); Kirshenbaum, *J. Clin. Invest.* 92:381-387 (1993); Roessler, *J. Clin. Invest.* 92:1085-1092 (1993); Moullier, *Nature Genetics* 4:154-159 (1993); La Salle, *Science* 259:988-990 (1993); Gomez-Foix, *J. Biol. Chem.* 267:25129-25134 (1992); Rich, *Human Gene Therapy* 4:461-476 (1993); Zabner, *Nature Genetics* 6:75-83 (1994); Guzman, *Circulation Research* 73:1201-1207 (1993); Bout, *Human Gene Therapy* 5:3-10 (1994); Zabner, *Cell* 75:207-216 (1993); Caillaud, *Eur. J. Neuroscience* 5:1287-1291 (1993); and Ragot, *J. Gen. Virology* 74:501-507 (1993)). Recombinant adenoviruses achieve gene transduction by binding to specific cell surface receptors, after which the virus is internalized by receptor-mediated endocytosis, in the same manner as wild type or replication-defective adenovirus (Chardonnet and Dales, *Virology* 40:462-477 (1970); Brown and Burlingham, *J. Virology* 12:386-396 (1973); Svensson and Persson, *J. Virology* 55:442-449 (1985); Seth, et al., *J. Virol.* 51:650-655 (1984); Seth, et al., *Mol. Cell. Biol.* 4:1528-1533 (1984); Varga et al., *J. Virology* 65:6061-6070 (1991); Wickham et al., *Cell* 73:309-319 (1993)).

A preferred viral vector is one based on an adenovirus which has had the E1 gene removed and these virons are generated in a cell line such as the

human 293 cell line. In another preferred embodiment both the E1 and E3 genes are removed from the adenovirus genome.

Another type of viral vector is based on an adeno-associated virus (AAV). This defective parvovirus is a preferred vector because it can infect many cell types and is nonpathogenic to humans. AAV type vectors can transport about 4 to 5 kb and wild type AAV is known to stably insert into chromosome 19. Vectors which contain this site specific integration property are preferred. An especially preferred embodiment of this type of vector is the P4.1 C vector produced by Avigen, San Francisco, CA, which can contain the herpes simplex virus thymidine kinase gene, HSV-tk, and/or a marker gene, such as the gene encoding the green fluorescent protein, GFP.

The inserted genes in viral and retroviral usually contain promoters, and/or enhancers to help control the expression of the desired gene product. A promoter is generally a sequence or sequences of DNA that function when in a relatively fixed location in regard to the transcription start site. A promoter contains core elements required for basic interaction of RNA polymerase and transcription factors, and may contain upstream elements and response elements.

Viral Promoters and Enhancers: Preferred promoters controlling transcription from vectors in mammalian host cells may be obtained from various sources, for example, the genomes of viruses such as: polyoma, Simian Virus 40 (SV40), adenovirus, retroviruses, hepatitis-B virus and most preferably cytomegalovirus, or from heterologous mammalian promoters, e.g. beta actin promoter. The early and late promoters of the SV40 virus are conveniently obtained as an SV40 restriction fragment which also contains the SV40 viral origin of replication (Fiers et al., *Nature* 273:113 (1978)). The immediate early promoter of the human cytomegalovirus is conveniently obtained as a HindIII E restriction fragment (Greenway, P.J. et al., *Gene* 18:355-360 (1982)). Of course, promoters from the host cell or related species also are useful herein.

Enhancer generally refers to a sequence of DNA that functions at no fixed distance from the transcription start site and can be either 5' (Laimins

et al., *Proc. Natl. Acad. Sci.* 78:993 (1981)) or 3' (Lusky et al., *Mol. Cell Bio.* 3: 1108 (1983)) to the transcription unit. Furthermore, enhancers can be within an intron (Banerji, J.L. et al., *Cell* 33:729 (1983)) as well as within the coding sequence itself (Osborne et al., *Mol. Cell Bio.* 4:1293 (1984)). They are usually between 10 and 300 bp in length, and they function in cis. Enhancers function to increase transcription from nearby promoters. Enhancers also often contain response elements that mediate the regulation of transcription. Promoters can also contain response elements that mediate the regulation of transcription. Enhancers often determine the regulation of expression of a gene. While many enhancer sequences are now known from mammalian genes (globin, elastase, albumin, α-fetoprotein and insulin), typically one will use an enhancer from a eukaryotic cell virus. Preferred examples are the SV40 enhancer on the late side of the replication origin (bp 100-270), the cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin, and adenovirus enhancers.

The promotor and/or enhancer may be specifically activated either by light or specific chemical events which trigger their function. Systems can be regulated by reagents such as tetracycline and dexamethasone. There are also ways to enhance viral vector gene expression by exposure to irradiation, such as gamma irradiation, or alkylating chemotherapy drugs.

It is preferred that the promoter and/or enhancer region act as a constitutive promoter and/or enhancer to maximize expression of the region of the transcription unit to be transcribed. It is further preferred that the promoter and/or enhancer region be active in all eukaryotic cell types. A preferred promoter of this type is the CMV promoter (650 bases). Other preferred promoters are SV40 promoters, cytomegalovirus (full length promoter), and retroviral vector LTF.

It has been shown that all specific regulatory elements can be cloned and used to construct expression vectors that are selectively expressed in specific cell types such as melanoma cells. The glial fibrillary acetic protein (GFAP) promoter has been used to selectively express genes in cells of glial origin.

Expression vectors used in eukaryotic host cells (yeast, fungi, insect, plant, animal, human or nucleated cells) may also contain sequences necessary for the termination of transcription which may affect mRNA expression. These regions are transcribed as polyadenylated segments in the untranslated portion of the mRNA encoding tissue factor protein. The 3' untranslated regions also include transcription termination sites. It is also preferred that the transcribed units contain other standard sequences alone or in combination with the above sequences improve expression from, or stability of, the construct.

Markers: The viral vectors can include nucleic acid sequence encoding a marker product. This marker product is used to determine if the gene has been delivered to the cell and once delivered is being expressed. Preferred marker genes are the *E. Coli* lacZ gene which encodes -galactosidase and green fluorescent protein.

In some embodiments the marker may be a selectable marker. Examples of suitable selectable markers for mammalian cells are dihydrofolate reductase (DHFR), thymidine kinase, neomycin, neomycin analog G418, hydromycin, and puromycin. When such selectable markers are successfully transferred into a mammalian host cell, the transformed mammalian host cell can survive if placed under selective pressure. There are two widely used distinct categories of selective regimes. The first category is based on a cell's metabolism and the use of a mutant cell line which lacks the ability to grow independent of a supplemented media. Two examples are: CHO DHFR$^-$ cells and mouse LTK$^-$ cells. These cells lack the ability to grow without the addition of such nutrients as thymidine or hypoxanthine. Because these cells lack certain genes necessary for a complete nucleotide synthesis pathway, they cannot survive unless the missing nucleotides are provided in a supplemented media. An alternative to supplementing the media is to introduce an intact DHFR or TK gene into cells lacking the respective genes, thus altering their growth requirements. Individual cells which were not transformed with the DHFR or TK gene will not be capable of survival in non-supplemented media.

The second category is dominant selection which refers to a selection scheme used in any cell type and does not require the use of a mutant cell line. These schemes typically use a drug to arrest growth of a host cell. Those cells which have a novel gene would express a protein conveying drug resistance and would survive the selection. Examples of such dominant selection use the drugs neomycin, (Southern and Berg, *J. Molec. Appl. Genet.* 1: 327 (1982)), mycophenolic acid, (Mulligan and Berg, *Science* 209: 1422 (1980)) or hygromycin, (Sugden et al., *Mol. Cell. Biol.* 5: 410-413 (1985)). The three examples employ bacterial genes under eukaryotic control to convey resistance to the appropriate drug G418 or neomycin (geneticin), xgpt (mycophenolic acid) or hygromycin, respectively. Others include the neomycin analog G418 and puramycin.

b. Carriers

The disclosed variant RNase P RNA molecules, and vectors encoding variant RNase P RNA, can also be delivered to cells using carriers. Numerous carriers are known for the delivery of compounds to cells and can be used for in RNase P compositions. RNase P RNA compositions may be, for example, in solution or suspension (for example, incorporated into microparticles, liposomes, or cells). These may be targeted to a particular cell type via antibodies, receptors, or receptor ligands. The following references are examples of the use of this technology to target specific proteins to tumor tissue (Senter et al., *Bioconjugate Chem.* 2:447-451 (1991); Bagshawe *Br. J. Cancer* 60:275-281, (1989); Bagshawe et al., *Br. J. Cancer* 58:700-703 (1988); Senter, et al., *Bioconjugate Chem.* 4:3-9 (1993); Battelli et al., *Cancer Immunol. Immunother.* 35:421-425 (1992); Pietersz and McKenzie, *Immunolog. Reviews* 129:57-80 (1992); and Roffler et al., *Biochem. Pharmacol* 42:2062-2065 (1991)). Vehicles such as "stealth" and other antibody conjugated liposomes (including lipid mediated drug targeting to colonic carcinoma), receptor mediated targeting of DNA through cell specific ligands, lymphocyte directed tumor targeting, and highly specific therapeutic retroviral targeting of murine glioma cells *in vivo*. The following references are examples of the use of this technology to target specific

proteins to tumor tissue (Hughes et al., *Cancer Research* 49:6214-6220 (1989); and Litzinger and Huang, *Biochimica et Biophysica Acta*, 1104:179-187 (1992)).

In general, receptors are involved in pathways of endocytosis, either constitutive or ligand induced. These receptors cluster in clathrin-coated pits, enter the cell via clathrin-coated vesicles, pass through an acidified endosome in which the receptors are sorted, and then either recycle to the cell surface, become stored intracellularly, or are degraded in lysosomes. The internalization pathways serve a variety of functions, such as nutrient uptake, removal of activated proteins, clearance of macromolecules, opportunistic entry of viruses and toxins, dissociation and degradation of ligand, and receptor-level regulation. Many receptors follow more than one intracellular pathway, depending on the cell type, receptor concentration, type of ligand, ligand valency, and ligand concentration. Molecular and cellular mechanisms of receptor-mediated endocytosis has been reviewed (Brown and Greene, *DNA and Cell Biology* 10(6):399-409 (1991)).

The disclosed variant RNase P RNA molecules, and vectors encoding variant RNase P RNA, can be used therapeutically in combination with a pharmaceutically acceptable carrier. Pharmaceutical carriers are known to those skilled in the art. Variant RNase P RNA molecules, and vectors encoding variant RNase P RNA, intended for pharmaceutical delivery may be formulated in a pharmaceutical composition. Pharmaceutical compositions may include carriers, thickeners, diluents, buffers, preservatives, surface active agents and the like in addition to the molecule of choice. Pharmaceutical compositions may also include one or more active ingredients such as antimicrobial agents, antiinflammatory agents, anesthetics, and the like.

The pharmaceutical composition may be administered in a number of ways depending on whether local or systemic treatment is desired, and on the area to be treated. Administration may be topically (including ophthalmically, vaginally, rectally, intranasally), orally, by inhalation, or parenterally, for example by intravenous drip, subcutaneous, intraperitoneal or intramuscular injection. The disclosed antibodies can be administered intravenously,

intraperitoneally, intramuscularly, subcutaneously, intracavity, or transdermally.

Preparations for parenteral administration include sterile aqueous or non-aqueous solutions, suspensions, and emulsions. Examples of non-aqueous solvents are propylene glycol, polyethylene glycol, vegetable oils such as olive oil, and injectable organic esters such as ethyl oleate. Aqueous carriers include water, alcoholic/aqueous solutions, emulsions or suspensions, including saline and buffered media. Parenteral vehicles include sodium chloride solution, Ringer's dextrose, dextrose and sodium chloride, lactated Ringer's, or fixed oils. Intravenous vehicles include fluid and nutrient replenishers, electrolyte replenishers (such as those based on Ringer's dextrose), and the like. Preservatives and other additives may also be present such as, for example, antimicrobials, anti-oxidants, chelating agents, and inert gases and the like.

Formulations for topical administration may include ointments, lotions, creams, gels, drops, suppositories, sprays, liquids and powders. Conventional pharmaceutical carriers, aqueous, powder or oily bases, thickeners and the like may be necessary or desirable.

Compositions for oral administration include powders or granules, suspensions or solutions in water or non-aqueous media, capsules, sachets, or tablets. Thickeners, flavorings, diluents, emulsifiers, dispersing aids or binders may be desirable. Formulations for parenteral administration may include sterile aqueous solutions which may also contain buffers, diluents and other suitable additives.

Dosing is dependent on the effect desired, but will normally be one or more doses per day, with course of treatment lasting from several days to several months or until one of ordinary skill in the art determines the delivery should cease. Persons of ordinary skill can easily determine optimum dosages, dosing methodologies and repetition rates.

The present invention will be further understood by reference to the following non-limiting examples.

Examples

Example 1: Evolving M1 RNA derivatives for DNA cleavage

This example describes generation of DNA-cleaving RNase P variants by *cis-trans in vitro* evolution.

A. Materials and Methods

1. Materials

Unlabeled nucleoside triphosphates (NTPs) and deoxynucleoside triphosphates (dNTPs) were purchased from Pharmacia. The radio-labeled chemicals [-^{32}P]ATP and [-^{32}-P]ATP were purchased from Amersham. Synthetic RNA and DNA oligonucleotides were obtained from the Keck Oligo-synthesis Center at Yale. The enzymes bacteriophage T4 polynucleotide kinase, Moloney murine leukemia virus (MoMLV) reverse transcriptase, and bacteriophage T7 RNA polymerase were obtained from New England Biolabs (NEB). TaqDNA polymerase, DNase I, and ribosomal RNA were obtained from Boehringer Mannheim. Streptavidin-coated paramagnetic beads were obtained from Dynal. Isolation of individual ribozymes was achieved by shotgun cloning of PCR product using a TOPO-TA cloning kit from Invitrogen.

2. Constructs and preparation of RNA

The M1GS construct used was derived from the pM1N2 plasmid provided by Dr. Sidney Altman and Ying Li (Li et al., *Proc Natl Acad Sci U S A* 89(8):3185-9 (1992)). Ribozymes were transcribed from DNA template generated by PCR amplification of the pM1N2 plasmid as previously described (Hanczyc & Dorit, *RNA* 4:268-275 (1998)). The forward primer contains a T7 RNA polymerase promoter (5'-CGAAATTAATACGACTCACTATAGAAGCTGACCAGACAG-3', SEQ ID NO:4), while the reserve primer anneals to the tethered guide sequence (5'-CTGGTGACAGGAGAATCCG-3', SEQ ID NO:5). Transcription reactions were performed as recommended by NEB and terminated by the addition of 2.5-times the reaction volume of ethanol. Precipitated reactions were resuspended in DNase I buffer, and DNA template was digested by

incubating 100 U DNase I for 40 minutes at 37°C. Transcripts were then gel purified on 5% polyacrylamide gels containing 8M urea.

3. Selection for DNA cleavage

A heterogeneous population of M1GS ribozymes was construct by doing seven consecutive amplifications of the pM1N2 plasmid using mutagenic PCR (Cadwell & Joyce, *PCR Methods Appl* 3(6), 136-140 (1994)). This PCR product was used as a template for a runoff transcription reaction which was followed by a DNase digestion of the DNA template. RNA transcripts were subsequently purified on a 10% acrylamide/8M Urea gel. The M1GS ribozyme variants were annealed to 5'biotinylated DNA oligonucleotide substrate (5'-Biotin-TCCTAACTGCTAACTGACAGGAGAATCCGA-3', SEQ ID NO:3) by adding 10-20 pmol M1GS to 40 pmol DNA substrate in annealing buffer (50 mM Tris (pH 8), 200 mM NH_4Cl). The mixture was heated to 65°C for 5 minutes, then slowly cooled to room temperature. The M1GS/DNA mixture was then combined with 20 l of Dynal streptavidin-coated paramagnetic beads (10^8 beads/ml) that were prepared as recommended by the manufacturer. To reduce non-specific binding of the M1GS/DNA complex, the beads were pre-incubated with 5 g of ribosomal RNA in annealing buffer. The M1GS/DNA complex was mixed with the beads by pipetting and allowed to bind for five minutes. The beads were then pelleted using a strong magnet and washed extensively by resuspension in annealing buffer. Cleavage of the DNA substrate was induced by adding a buffer containing 50 mM Tris (pH 8.0), 200 mM NH_4Cl and either 20 mM or 100 mM $MgCl_2$. Cleavage reactions were incubated at 24°C for a time ranging from 3 hours in the first generation of selection, to 10 minutes in the twenty-fifth generation. Ribozymes recovered in the cleavage reaction supernatant were reverse transcribed using MoMLV. cDNAs were amplified by either standard or mutagenic PCR and used a sDNA template for transcription of the next generation.

4. Kinetic assays

Activity of M1 RNA variants was measured by ^{32}P-labeling of the appropriate substrates and quantification of the cleavage products using a Fuji BAS 2000 phosphorimaging system. The DNA and RNA oligonucleotide substrates were 5'-labeled while the ptRNA and p4.5S RNA substrates were internally labeled in transcription. The parameters K_M, k_{cat} and k_{obs}, were determined using single-turnover kinetics (Herschlag & Cech, *Biochemistry* 29(44):10159-10171 (1990); McConnell et al., *Proc Natl Acad Sci U S A* 90(18):8362-8366 (1993); Tsang & Joyce, *J Mol Biol* 262(1):31-42 (1996)). In short, cleavage reactions contained [E]>>[S] (30 nM – 1.5 uM enzyme and 1-2 nM substrate), and four time points were taken over a range of enzyme concentrations that spanned the K_M. Ribozymes were pre-incubated in 5 mM $MgCl_2$ for 5 minutes before being added to the reaction. Reactions were incubated at 37°C. The k_{cat} and K_M values were determined individually by plotting k_{obs} over a range of ribozyme concentrations that spanned K_M. Data were fit to the equation $k_{obs} = k_{cat}/([E]+ K_M)$, which holds true when $[S] << K_M$, assuming that binding of the ribozyme and substrate has reached pre-equilibrium.

5. Cleavage assays in *trans*

Activity of the *trans* substrates (ptRNA, p4.5S RNA, p4.5S RNA and DNA/EGS) was determined using natural length (377 nt) ribozymes lacking tethered guide sequences. This ensured that the tethered guide sequences of M1GS, G6-3 and G25-10 did not interfere with the binding or cleavage of *trans* substrates. In addition to ptRNA and p4.5S RNA, DNA cleavage was assayed in *trans* using external RNA guide sequences to target the DNA oligonucleotide (Forster & Altman, *Science* 249(4970), 783-786 (1990); Li & Altman, *Nucleic Acids Res.* 24(5):835-842 (1996)). External RNA guide sequences (EGSs) used in this analysis included the tether and base pairing guide sequence from the M1GS construct (Figure 1A and Figure 4). The RNA EGSs and DNA substrate were annealed by combining 100 pmol of each primer in 20 l of annealing buffer, heating to 65°C for 5 minutes and slow cooling. Clones G6-3 and G25-10 contain mutations in their guide

sequence tethers and were assayed with separate external guide sequences containing these mutations. The sequences containing these mutations. The sequences for RNA external guide sequences are:

M1 RNA wild-type EGS (nucleotides 378 to 417 of SEQ ID NO:1):

5'-GAUUUACGUCAUCCACUCGAUCGGAUUCUCCUGUCACCAG-3'

G6-3 EGS (SEQ ID NO:6):

5'-GAUUUACGUCAUCCACUCAAUCGGAUUCUCCUGUCACCAG-3'

G25-10 EGS (SEQ ID NO:7):

5'-GAUUUACGGCAUCACUCAAUCGGAUUCUCCUGUCACCAG-3'

6. Sequence Analysis

Individual ribozymes were isolated by shotgun cloning PCR product using a Invitrogen TOPO-TA cloning kit. The forward and reverse strands of ribozyme-containing plasmids were sequenced on an automated ABI 373 sequencer. Fifty clones were sequenced from generation six, forty from generation twenty-five and ten from the initial starting pool G0.

7. Construction of point mutants

Mutations were made at specific positions in the ribozyme using a Quick-change site-directed mutagenesis kit from Stratagene. The following primers were used:

Construction of M1U59:

M1U59.fwd (SEQ ID NO:8)

5'-GACGGGCGGAGTGGAGGAAAGTCC-3'

M1U59.rev (SEQ ID NO:9)

5'-GACGGGCGGAGGGGAGGAAAGTCC-3'

M1G59.fwd (SEQ ID NO:10)

5'-GACGGGGCGGAGGGGAGGAAAGTCC-3'

M1G59.rev (SEQ ID NO:11)

5'-GGACTTTCCTCCCCTCCGCCCGTC-3'

M1G173.fwd (SEQ ID NO:12)

5'-GGGATCAGGTAAGGGTGAAAG-3'

M1G173.rev (SEQ ID NO:13)

5'-CTTTCACCCTTACCTGATCCC-3'

M1U228.fwd (SEQ ID NO:14)
5'-GTCCGTGGCATGGTAAACTCC-3'
M1U228.rev (SEQ ID NO:15)
5'-GGAGTTTACCATGCCACGGAC-3'
M1C228.fwd (SEQ ID NO:16)
5'-GTCCGTGGCACGGTAAACTCC-3'
M1C228.rev (SEQ ID NO:17)
5'-GGAGTTTACCGTGCCACGGAC-3'
M1C299.fwd (SEQ ID NO:18)
5'-CCCGGGTAGGCCGCTTGAGCCAGTG-3'
M1C299.reV (SEQ ID NO:19)
5'-CACTGGCTCAAGCGGCCTACCCGGG-3'
M1396A to G (SEQ ID NO:20):
5'-CTGGTGACAGGAGAATCCGATCGAGTGGATGACGTAAATC-3'

B. Evolving M1 RNA derivatives for DNA cleavage

Selection for enhanced DNA cleavage was performed using an M1 RNA construct (M1GS) that contains a guide sequence tethered to the 3'-end of the ribozyme (Figure 1A)(Li et al., *Proc Natl Acad Sci U S A* 89(8):3185-9 (1992); Liu & Altman, *Genes Dev* 9(4):471-480 (1995); Pace & Brown, *J Bateriol* 177(8):1919-1928 (1995)). The guide sequence base-pairs with a complementary DNA oligonucleotide substrate forming a helical stem that is cleaved M1 RNA (Forster & Altman, *Science* 249(4970), 783-786 (1990)). A 3' ACCA motif in the guide sequence directs the cleavage site which lies 5' of the M1GS/Substrate helix. Figure 1B outlines the selection procedure.

A heterogenous population of 10^{13} M1GS molecules was created using seven consecutive rounds of mutagenic PCR (Cadwell & Joyce, *PCR Methods Appl* 3(6), 136-140 (1994)). This population was initially selected for five generations under permissive conditions containing a high divalent ion concentration (100 mM $MgCl_2$). Optimal M1 RNA activity *in vitro* requires high concentrations of divalent ions (e.g., Mg^{2+}) which provide electrostatic shielding and are involved in the mechanism of catalysis (Guerrier-Takada et al., *Biochemistry* 25(7):1509-15 (1986); Perreault &

Altman, *J. Mol. Biol.* 230:750-756 (1993); Smith & Pace, *Biochemistry* 32(20):5273-81 (1993)). Selection under permissive conditions was necessary to enrich the set with active variants because most molecules in the mutagenized starting population were unable to perform DNA cleavage. Selection was continued after the first five generations, under more stringent conditions (20 mM $MgCl_2$), for a total of twenty-five generations.

DNA cleavage activity of the population steadily increased during the selection procedure. Activity of the highly mutagenized starting population was undetectable, reflecting the deleterious effect of most mutations. The rate of DNA cleavage (k_{obs}) increased to 100-times that of wild-type M1GS following the initial five rounds of selection. Activity continued to increase during the next twenty generations, eventually resulting in a population activity 1000-fold higher than that of wild-type M1GS. The rate of improvement in DNA cleavage declined significantly after generation twenty-three, suggesting that the population has reached an activity plateau.

Example 2: Sequence Analysis

To track the process of molecular adaptation at the genotypic level, ribozymes were sequenced at three stages of selection: the initial population (G0), the population after six generations of selection (G6), and after twenty-five generations of selection (G25). Sequences from the mutagenized initial population (G0) provide a glimpse of the variability present in the starting pool. Generation six follows the first appearance of significant levels of DNA cleavage; by generation twenty-five the population had reached an activity plateau. A comparison of sequences from these three time points was meant to uncover mutations associated with increasing DNA cleavage.

Sequences from the initial population (G0) contained in average of 18.3 mutations per molecule compared to wild-type M1GS, with a range of 14-24 mutations. These mutations were randomly distributed throughout the molecule (based on a G-test for goodness of fit, block analysis p-value >0.1) with no apparent mutational bias (Sokol & Rohlf, *Biometry*. Second Edition, W. H. Freeman and Company (1981)). Generation six (G6) displayed a marked decrease in the average number of mutations per molecule

(mean=11.2 mutations/molecule; range=7-16), the result of strong selection for improved cleavage activity. Mutagenic PCR was performed after generations 6, 11 and 17 to restore population variability. Generation twenty-five (G25) sequences displayed an increased number of mutations (mean=17.2 mutations/molecule; range=15-21), reflecting an accumulation of mutations during the course of evolution.

Fixed or prevalent mutations have been used as a starting point in identifying genotypic changes important for conferring an evolved phenotype (Beaudry & Joyce, *Science* 257(5070):635-641 (1992); Frank & Pace, *Proc Natl Acad Sci U S A* 94(26):14355-14360 (1997); Tsang & Joyce, *J Mol Biol* 262(1):31-42 (1996)). The simple rationale underlying these analyses is that intense directional selection should result in the rapid fixation of mutations that confer an increase in activity. Sequence analysis of fifty G6 clones revealed two prominent mutations in the population: a 59:G to U change appears in all 50 of the 50 clones and a 299:U to C change appears in 48 of 50 clones. Generation twenty-five sequences contain three fixed mutations (40 of 40 sequences) at position 73 (73:A to G), position 288 (288:C to U), and position 299 (299:U to C). The 59:G to U mutation, ubiquitous in generation 6, is no longer fixed: 4 of 40 sequences contain a new mutation, 59:G to C, while 36 of 40 retain the 59:G to U change.

Figure 2 summarizes the frequency and distribution of mutations from the G6 and G25 populations. Note that all three of the mutations that are fixed in generation 25 are present in at least 30% of the G6 sequences. Three general motifs are apparent from the G25 sequence data. First, there are no mutations in the P4 pseudo-knot; which is thought to be part of the catalytic core and is known to be important for catalysis (Haas et al., *Proc Natl Acad Sci U S A* 91(7):2527-31 (1994)). Second, the P1 stem accumulates mutations near the 5'-end (positions 368-373) that disrupts base pairing and likely increase the flexibility of the guide sequence tether (see Figure 1A). Third, mutations that preserve base pairing in the P3 stem accumulate at positions 20-38, suggesting that structure, and not sequence identity, is of primary importance to the function of this region.

Example 3: Individual Clones

Two representative clones were chosen to investigate the kinetic basis for improved DNA cleavage, G6-3 and G25-10 (see Figure 3). These clones were chosen based on two criteria: 1) they are genotypically characteristic, and 2) they display cleavage activity comparable to that of the populations from which they are drawn.

Product release is known to be the rate-limiting step in M1 RNA processing of pre-tRNA (Tallsjo & Kirsebom, *Nucleic Acids Res* 21(1):51-57 (1993)) and is likely to be the kinetically limiting step in this *in vitro* selection system (Herschlag, *Proc Natl Acad Sci U S A* 88(16):6921-6925 (1991); Herschlag & Cech, *Biochemistry* 29(44):10159-10171 (1990)). Therefore, kinetic assays were performed under single-turnover conditions to largely eliminate the product-release component of the first-order rate constant. Under these conditions the constant k_{cat} is primarily a measure of the actual DNA cleavage step.

The kinetic analyses shown in Table 1 indicate that enhanced DNA cleavage in G6-3 and G25-10 is attributable to an increased k_{cat}, suggesting an accelerated cleavage step. Despite the dramatic improvement in DNA cleavage, the overall catalytic efficiency (k_{cat}/K_M) of these evolved molecules is offset by reduced substrate binding as measured by K_M.

Table 1: Kinetic characterization of DNA cleavage under stringent selection conditions (20 mM $MgCl_2$).

	$k_{cat}(min^{-1})$	K_M(nM)	k_{cat}/K_M ($min^{-1}M^{-1}$)
M1GS (wt) x10^2	$1.3^{\pm}0.6 \times 10^{-6}$	$50^{\pm}8$	$2.5^{\pm}0.8$
G6-3 x10^4	$3.2^{\pm}0.2 \times 10^{-4}$	$94^{\pm}11$	$3.4^{\pm}0.2$
G25-10 x10^5	$9.5^{\pm}0.5 \times 10^{-3}$	$208^{\pm}13$	$4.6^{\pm}0.2$

Example 4: DNA Cleavage in *trans*

The M1GS construct used in the selection differs from wild-type M1RNA by the addition of a guide sequence linked to the 3' end of the molecule. The DNA substrate was supplied in *trans*. This was accomplished by pre-annealing DNA oligonucleotide to an external (untethered) RNA guide sequence identical to the M1GS tether and guide sequence (see Figure 4). Cleavage of this hybrid double-stranded helix was then assayed in *trans* using native length (377-nt) M1RNA and evolved ribozymes lacking the tethered guide sequence. Kinetic analysis indicated that the enhanced DNA cleavage exhibited by G6-3 and G25-10 does not depend on a tethered substrate. Although catalysis in *trans* is less efficient less efficient than in *cis*, the relative improvements of G6-3 and G25-10 over wild-type are accentuated in *trans*. For example, under permissive cleavage conditions (100 mM $MgCl_2$) clone G6-3 shows a modest 10-fold improvement in *cis* DNA cleavage efficiency (k_{cat}/K_M) over wild-type M1 RNA (See Table 2). In *trans*, however, this improvement increases to 30-fold greater than wild-type. Under stringent selection conditions (20 mM $MgCl_2$), the activity of the wild-type ribozyme on a DNA substrate supplied in *trans* is undetectable, while G25-10 retains appreciable activity. These results indicate that evolved DNA cleavage activity is not simply an adaptation to a tethered substrate.

Table 2: Cleavage efficiency (k_{cat}/K_M ($min^{-1}M^{-1}$)) in *trans* and *cis* on a DNA oligonucleotide at 20 mM and 100 mM $MgCl_2$.

	cis 100 mM	*trans* 100 mM	*cis* 20 mM	*trans* 20 mM
M1 RNA	$6.5^{\pm}1.0 \times 10^4$	$1.5^{\pm}0.5 \times 10^3$	$2.5^{\pm}0.8 \times 10^2$	n.d.
G6-3	$7.2^{\pm}1.1 \times 10^5$	$4.8^{\pm}1.5 \times 10^4$	$3.4^{\pm}0.2 \times 10^4$	n.d.
G25-10	$4.5^{\pm}0.6 \times 10^6$	$1.6^{\pm}0.2 \times 10^5$	$4.6^{\pm}0.2 \times 10^5$	$1.8^{\pm}0.5 \times 10^4$

Cleavage assays performed in *trans* used native length, 377-nt ribozymes to ensure that the tethered guide sequence did not interfere with catalysis.
All *cis* reactions used ribozymes containing a 3' tethered guide sequence.
n.d.= non detectable.

Example 5: Probing the DNA cleavage phenotype: analysis of point mutations

The examination of ribozyme sequences from generations 6 and 25 highlights mutations that appear in conjunction with increased DNA cleavage. Four prevalent mutations were tested for their relative contribution to enhanced DNA cleavage: 59:G to U, 173:A to G, 228:C to U and 299:U to C. These mutations were placed individually in a wild-type M1GS background and were also reverted to the corresponding wild-type nucleotide in the clone G25-10 background. None of the four mutations, when placed alone in a wild-type M1GS background, significantly improved DNA cleavage. Conversely, three of the four G25-10 revertants did not show diminished DNA cleavage: reverting 228:U to wild-type in clone G25-10 causes a 4-fold reduction in DNA cleavage by affecting both k_{cat} and K_M. Thus, multiple mutations likely have a collective effect on DNA-cleavage activity.

Clones G6-3 and G25-10 also contain a 370:U to G mutation in stem P1. Although this mutation appears in less than 70% of the G6 and G25 sequences, it was chosen for examination because phylogenetic comparisons suggest that base pairs 3:A::371:U and 4:G::370:U are part of a long range tertiary interaction with tetraloop L9 (Massire et al., *RNA* 3(6):553-556 (1997)). The 370:U to G mutation might therefore alter this interaction and affect DNA cleavage. Reverting this position to the wild-type sequence in both clones G6-3 and G25-10 did not, however, diminish DNA cleavage.

In contrast, a mutation in the single-stranded tether (396:G to A) linking the guide sequence to the ribozyme does affect DNA cleavage (see Table 3). Although this 396:G to A mutation appears in less than 50% of the generation 25 sequences, this mutation was found to be functionally significant when placed in a wild-type M1GS background (M1GS.396A) or when reverted to the wild-type sequence in both representative clones (G6-3.396G and G25-10.396G). The construct M1GS.396A shows a 4-fold increase in DNA cleavage due to an improvement in both k_{cat} and K_M.

Reverting the mutation in the evolved clones resulted in a 7-fold reduction in activity for G6-3, and a 10-fold reduction in activity for G25-10.

Table 3: Kinetic characterization of point mutations affecting DNA cleavage.

	$k_{cat}(min^{-1})$	$K_M(nM)$	k_{cat}/K_M $(min^{-1}M^{-1})$
M1GS (wt)	$1.3^{\pm}0.6 \times 10^{-6}$	$50^{\pm}8$	$2.5^{\pm}0.8 \times 10^{2}$
M1GS.396A	$3.5^{\pm}0.9 \times 10^{-6}$	$35^{\pm}5$	$9.8^{\pm}1.1 \times 10^{2}$
G6-3	$3.2^{\pm}0.2 \times 10^{-4}$	$94^{\pm}11$	$3.4^{\pm}0.2 \times 10^{4}$
G6-3.396G	$1.5^{\pm}0.3 \times 10^{-4}$	$300^{\pm}25$	$5.0^{\pm}0.6 \times 10^{3}$
G25-10	$9.5^{\pm}0.5 \times 10^{-3}$	$208^{\pm}13$	$4.6^{\pm}0.2 \times 10^{5}$
G25-10.228C	$3.7^{\pm}0.7 \times 10^{-3}$	$290^{\pm}22$	$1.2^{\pm}0.2 \times 10^{5}$
G25-10.396G	$2.2^{\pm}0.2 \times 10^{-3}$	$507^{\pm}31$	$4.3^{\pm}0.2 \times 10^{4}$
G25-10.228C396G	$2.0^{\pm}0.2 \times 10^{-3}$	$520^{\pm}24$	$3.8^{\pm}0.2 \times 10^{4}$

These results, along with the population changes described in Example 2, indicate that multiple mutations in combination produce the observed efficient DNA-cleaving activity.

Example 6: Substrate versatility and the activity of evolved ribozymes on RNA substrates

The substrate versatility of the evolved ribozymes was examined by assaying clones G6-3 and G25-10 on three RNA substrates: an RNA oligonucleotide analog of the DNA substrate and two substrates that are processed by *E. coli* RNase P *in vivo* – pre-$tRNA^{Tyr}$ and pre-4.5S RNA. As shown in Figure 5, both G6-3 and G25-10 are unable to process the ptRNA substrate efficiently, showing, respectively, a 5000-fold and 300-fold reduction in efficiency (k_{cat}/K_M) compared to wild-type M1 RNA. The G6 and G25 populations also show a reduction in $ptRNA^{Tyr}$ cleavage comparable to that of clones C6-3 and G25-10. These results show that the $ptRNA^{Tyr}$ cleavage deficiency in not restricted to the representative clones but is a characteristic of the populations as a whole.

Evolved ribozymes assayed for the cleavage of the RNA oligonucleotide and p4.5S RNA reveal a more complex activity pattern.

Clone G6-3 exhibits lower than wild-type efficiency (k_{cat}/K_M) on all three RNA substrates tested: a 5000-fold reduction on ptRNA, a 5-fold reduction on the RNA oligonucleotide and a 2-fold reduction on p4.5S RNA (see Table 4 for cleavage of RNA substrates).

Table 4: RNA processing kinetic properties of wild-type and evolved ribozymes.

	$k_{cat}(min^{-1})$	$K_M(nM)$	k_{cat}/K_M ($min^{-1}M^{-1}$)
RNA Oligonucleotide			
M1GS (wt)	$1.2^{\pm}0.1 \times 10^{-2}$	$62^{\pm}4$	$1.9^{\pm}0.03 \times 10^{6}$
G6-3	$4.4^{\pm}0.4 \times 10^{-3}$	$105^{\pm}9$	$4.2^{\pm}0.02 \times 10^{5}$
G25-10	$4.3^{\pm}0.3 \times 10^{-2}$	$229^{\pm}21$	$1.9^{\pm}0.03 \times 10^{6}$
ptRNA			
M1 RNA (wt)	$3.4^{\pm}0.4 \times 10^{-1}$	$5478^{\pm}480$	$6.0^{\pm}0.4 \times 10^{5}$
G6-3	n.d.	n.d.	$1.2^{\pm}0.3 \times 10^{2}$
G25-10	$5.0^{\pm}1.4 \times 10^{-4}$	$2652^{\pm}354$	$1.9^{\pm}0.3 \times 10^{3}$
p4.5S RNA			
M1 RNA (wt)	n.d.	n.d.	$1.4^{\pm}0.2 \times 10^{4}$
G6-3	n.d.	n.d.	$5.6^{\pm}0.4 \times 10^{3}$
G25-10	n.d.	n.d.	$4.7^{\pm}0.2 \times 10^{4}$

Cleavage assays performed in *trans* (e.g. ptRNA andp4.5S RNA) used native length, 377-nt ribozymes to ensure that the tethered guide sequence did not interfere with catalysis.
All *cis* reactions (e.g., DNA and RNA oligonucleotides) used ribozymes containing a tethered guide sequence.
n.d.= not determined

By comparison, G25-10 cleaves ptRNA 300-fold less efficiently than wild-type, while it cleaves the RNA oligonucleotide with wild-type efficiently and cleaves p4.5S RNA 4-fold more efficiently than wild-type (see Figure 5). These results suggest that continued selection for DNA cleavage reverses the trend toward a net decline in activity on all RNA substrates seen at generation 6. By generation twenty-five, the correlation between improved DNA cleavage and the recovery of activity on the RNA

oligonucleotide and p4.5S RNA suggests that G25-10 processes these substrates similarly.

Although the site of DNA oligonucleotide cleavage remained constant during the course of *in vitro* evolution, clone G25-10 exhibits aberrant cleavage behavior on a fraction of the RNA oligonucleotide substrate. Cleavage reactions of the RNA oligonucleotide lasting longer than one substrate half-life (e.g., 30 minutes) resulted in the appearance of a 14-nt cleavage product in addition to the standard 15-nt product. This 14-nt product was produced by a one nucleotide shift of the cleavage site toward the 5'-end of the substrate binding exhibited by G25-10 (see Table 4) suggests an altered positioning of the substrate in the evolved clone.

Example 7: Measuring substrate specificity

The disclosed *in vitro* evolutionary system allows changes in the substrate specificity of M1 RNA derivatives to be observed as they evolve cleavage of a novel DNA substrate. Substrate specificity can be measured as the ratio of the kinetic specificity constants for DNA and RNA cleavage ($[k_{cat}/K_M DNA] / [k_{cat}/K_M RNA]$). This ratio measures DNA specificity with respect to each RNA substrate: the larger the value, the greater the preference for a DNA substrate. Table 5 illustrates the relative DNA substrate specificity of wild-type M1 RNA, G6-3 and G25-10. The low values for the wild-type ribozyme underscore its preference for all three RNA substrates over the DNA substrate. By comparison, G6-3 and G25-10 show approximately a 250-fold preference for the DNA substrate over ptRNA, and a 6 to 10-fold preference over p4.5S RNA. These preferences represent a 10^2 and 10^5 increase in DNA substrate specificity relative to the wild-type substrate profile. Finally, evolved clones continue to show a slight preference for the RNA oligonucleotide over its DNA counterpart (i.e., ratio<1). This preference is, however, significantly reduced from that of wild-type, suggesting an ever increasing DNA substrate specificity.

Table 5: Relative substrate preferences of wild-type and evolved ribozymes.

	RNA oligonucleotide	p4.5S RNA	ptRNA
M1RNA	0.0001	0.018	0.0004
G6-3	0.08	6.1	283
G25-10	0.24	9.8	242

Note: Values are the ratio of DNA/RNA specificity constants (k_{cat}/K_M). Larger values denote a preference for the DNA substrate.

Although catalysis in *trans* can be less efficient than in *cis*, the relative improvements of G6-3 and G25-10 over wild-type are often accentuated in *trans*. For example, under permissive cleavage conditions (100 mM $MgCl_2$) clone G6-3 shows a modest 10-fold improvement in *cis* DNA cleavage efficiency (k_{cat}/K_M) over wild-type M1 RNA (see Table 2). In *trans*, however, this improvement increases to 30-fold greater than wild-type. Under stringent selection conditions (20 mM $MgCl_2$), the activity of the wild-type ribozyme on a DNA substrate supplied in *trans* is undetectable, while G25-10 retains appreciable activity. These results indicate that evolved DNA cleavage activity is not simply an adaptation to a tethered substrate.

Example 8: *In Vitro* Evolution in the Presence of RNase P Protein Subunit

A. Materials and Methods

1. Materials

Unlabeled nucleoside triphosphates (NTPs) and deoxynucleoside triphosphates (dNTPs) were purchased from Pharmacia. Radio-labeled chemicals [α-^{32}P]ATP and [γ-^{32}P]ATP were purchased from Amersham. Synthetic RNA and DNA oligonucleotides were obtained from the Keck Oligo-synthesis Center at Yale. The enzymes bacteriophage T4 polynucleotide kinase, Moloney murine leukemia virus (MoMLV) reverse transcriptase, and bacteriophage T7 RNA polymerase were obtained from

New England Biolabs (NEB). Taq DNA polymerase, DNase I, and ribosomal RNA were obtained from Boehringer Mannheim. Streptavidin-coated paramagnetic beads were obtained from Dynal. Isolation of individual ribozymes was achieved by shotgun cloning of PCR product using a TOPO-TA cloning kit from Invitrogen. CM Sephadex was from Pharmacia Biotech.

2. Constructs and Preparation of RNA

The M1K1 construct used in this study is derived from the pM1N2 plasmid provided by Dr. Sidney Altman and Ying Li [Li et al., *Proc Natl Acad Sci U S A* 89(8):3185-9 (1992)]. A PCR fragment containing the T7 RNA polymerase promoter fused to the 5'-terminus of M1 RNA was created by amplification of the pM1N2 plasmid using the following primers: M1T7 (5'-CGAAATTAATACGACTCACTATAGAAGCTGACCAGACAG-3', SEQ ID NO:21) and MC377 (5'-AGGTGAAACTGACCGATAAG-3', SEQ ID NO:22). A tether and guide sequence were then added to the 3'-terminus of M1 RNA by using the T7-M1 RNA PCR product as template in a reaction containing three primers: M1K1splint (5'-CTGGTGAGGCCGTGTCTGCGCAGGAGTGTCATGGATGGGGTAAATTAGGTGAAACTGACCGATAA-3', SEQ ID NO:23), M1T7 and M1K1RT (5'-CTGGTGAGGCCGTGTCTGCGC-3', SEQ ID NO:24). The resultant PCR product was then cloned into the vector pCR2.1 (Invitrogen) and verified by sequencing both strands. Ribozymes were transcribed from DNA template generated by PCR amplification of the pM1K1 plasmid using the primers M1K1RT and M1T7. Transcription reactions were performed as recommended by New England Biolabs and cleaned using Quiagen RNEasy columns. The cleaned reaction was eluted in 50 ul of water and the DNA template was digested with 60 units of DNase I for 40 minutes at 37^0C. Transcripts were then gel purified on 5% polyacrylamide gels containing 8M urea and eluted by passive diffusion.

The RNA substrates $ptRNA^{Tyr}$ and p4.5S RNA were prepared by digestion of plasmids encoding the natural *E. coli* sequences for these precursors. The plasmids were linearized by digestion with FokI and SmaI,

respectively, and 20 uCi of [α-^{32}P]ATP was added to the run-off transcription reaction.

3. Purification of C5 Protein

C5 protein was isolated following the procedure detailed by Gopalan et al., *J Mol Biol* 267(4):818-829 (1997). Briefly, BL21(DE3) cells were transformed with the plasmid pBSC5 containing wild-type C5. The cells were grown to an A_{600} =0.4 and then induced with 2 mM IPTG. The crude cell extract was prepared by sonication and C5 protein was concentrated by a series of centrifugations. The protein was further purified using CM Sephadex C50 chromatography. Protein concentration was determined using a Bradford assay and measuring the absorbance at 595 nM.

4. Selection for DNA Cleavage

A heterogeneous population of M1K1 ribozymes was construct by doing six consecutive mutagenic PCR amplifications starting with the pM1K1 plasmid [Cadwell and Joyce, "Mutagenic PCR," *PCR Methods Appl* 3(6):136-140 (1994)]. The final PCR product was used as a template for runoff transcription to create the starting population (G0) and was also TA-cloned for sequencing. To select for DNA cleavage, the M1K1 ribozyme variants were annealed to 5'-biotinylated DNA oligonucleotide substrate (5'-Biotin-TCCTACCTGCTAACTGAGGCCGTGTCTGCG-3', SEQ ID NO:25) by adding 10-20 pmol M1K1 to 40 pmol DNA substrate in binding buffer (20 mM Hepes-K (pH=8 at 25^{0}C), 400 mM NH_4OAc, 2 mM $Mg(OAc)_2$, 0.01% Nonidet P-40) in a volume of 10 µl. The mixture was heated to 65^0 C for 5 minutes, then cooled to room temperature over five minutes. Forty µl of binding buffer containing 100 nM C5 protein was then added to the holoenzyme replicates while forty µl of binding buffer was added to the RNA-only replicates. The reactions were incubated for 5 minutes at 37^{0}C then the mixtures were combined with 20 µl of Dynal streptavidin-coated paramagnetic beads (10^8 beads/ml) that were prepared as recommended by the manufacturer. To reduce non-specific binding of the M1K1/DNA complex, the beads were pre-incubated with 5 ug of ribosomal RNA in binding buffer. The M1K1/DNA complex was mixed with the

beads by pipetting and allowed to bind for five minutes. The beads were then pelleted using a strong magnet and washed extensively by resuspension in binding buffer. Cleavage of the DNA substrate was induced by adding a buffer containing 20 mM Hepes-K (pH = 8 at 25^{0}C), 400 mM NH_4OAc, 0.01% Nonidet P-40 and either 20 mM or 100 mM $Mg(OAc)_2$. Cleavage reactions were incubated at 24^0 C for a time ranging from 3 hrs in the first generation of selection, to 5 minutes in the tenth generation. Ribozymes recovered in the cleavage reaction supernatant were reverse transcribed using MoMLV. cDNAs were amplified by either standard or mutagenic PCR and used as DNA template for transcription of the next generation.

5. Kinetic Assays

Activity of M1 RNA variants was measured by ^{32}P-labeling of the appropriate substrates and quantitation of the cleavage products using a Fuji BAS 2000 phosphorimaging system. The DNA and RNA oligonucleotide substrates were 5'-labeled while the ptRNA and p4.5S RNA substrates were internally labeled in transcription. Single-turnover kinetics were performed as follows: cleavage reactions contained [E]>>[S] (1 nM – 1 μM enzyme and 0.1-1 nM substrate), four time points were taken over a range of enzyme concentrations that spanned the K_M. Ribozymes were pre-incubated in 5 mM $MgCl_2$ for 5 minutes before being added to the reaction. Kinetic assays in the presence of C5 protein were performed by first renaturing the ribozyme, then adding a 2-10 fold excess of C5 protein and incubating for 5' at 37^0 C before adding the substrate. Cleavage reactions were carried out in buffer containing 20 mM Hepes-K (pH=8 at 25^{0}C), 400 mM NH_4OAc, 0.01% Nonidet P-40 and either 20 mM or 100 mM $Mg(OAc)_2$. Reactions were incubated at 37^{0}C. The k_{cat} and K_M values were determined individually by plotting k_{obs} over a range of ribozyme concentrations that spanned K_M. Data were fit to the equation $k_{obs}=k_{cat}/([E]+K_M)$, which holds true when $[S]<<K_M$, assuming that binding of the ribozyme and substrate has reached pre-equilibrium [Herschlag and Cech, *Biochemistry* 29(44):10159-10171 (1990); McConnell et al., *Proc Natl Acad Sci U S A* 90(18):8362-8366 (1993); Tsang and Joyce, *J Mol Biol* 262(1):31-42 (1996)].

Activities on the substrates ptRNA and p4.5S RNA were determined using natural length (377 nt) ribozymes lacking tethered guide sequences. This ensured that the tethered guide sequences of M1K1, and the evolved clones did not interfere with the binding or cleavage of *trans* substrates.

6. Gel-retardation Analysis of RNA-protein Binding

The gel-retardation analysis used in this study to determine the M1 RNA-C5 protein dissociation constant (K_d) is based on the protocol detailed by Talbot and Altman, *Biochemistry* 33(6):1399-405 (1994). Transcripts of M1-RNA, M1K1 and evolved clones were internally labeled with α-^{32}P ATP, gel-purified and renatured in binding buffer (20 mM Hepes-K (pH=8 at 25^{0}C), 400 mM NH_4OAc, 2 mM $Mg(OAc)_2$, 0.01% Nonidet P-40) by heating to 65^{0}C for 5 minutes then slowly cooling to 37^{0}C. M1 RNA (1 nM) was incubated with C5 protein (0.1 nM-25 nM prepared by serial dilution) in binding buffer containing 5% glycerol (v/v) at 37^{0}C for 5 minutes. RNA-protein mixtures were then resolved on non-denaturing 4% acrylamide gels (29:1 acrylamide: bis-acrylamide) containing 20 mM Hepes-K pH8, 1 mM $MgCl_2$. Gels were run at 30 mA for 9-12 hours at 4 ^{0}C with constant recirculation of the running buffer (20 mM Hepes-K pH8, 1 mM $MgCl_2$). Gels were quantified using a Fuji BAS 2000 phosphorimaging system and K_d was determined from plotting the fraction of RNA bound vs. C5 concentration where K_d=[C5][RNA]/[C5RNA].

7. Sequence Analysis

Individual ribozymes were isolated by shotgun cloning of PCR product using a Invitrogen TOPO-TA cloning kit. The forward and reverse strands of ribozyme-containing plasmids were sequenced on an automated ABI 373 sequencer. Fifty clones were sequenced from generation six, forty from generation twenty-five and ten from the initial starting pool, G0.

8. Construction of Point Mutants

Mutations were made at specific positions in the ribozyme using a Quick-change site-directed mutagenesis kit from Stratagene. The following primers were used:

A333 to U:

5'-GGCCTAGATGTATGACTGTCC-3' (SEQ ID NO:26)

5'-GGACAGTCATACATCTAGGCC-3' (SEQ ID NO:27)

U136 to A

5'-CAGTGCAACAGAGAGCAAACCGCCGATG-3' (SEQ ID NO:28)

5'-CATCGGCCGGTTTGCTCTCTGTTGCACTG-3' (SEQ ID NO:29)

B. Results

1. M1 RNA Evolves DNA Cleavage in the Presence of C5 Protein

The examples above show that derivatives of M1 RNA containing a tethered guide sequence (M1GS) could be evolved *in vitro* for efficient DNA cleavage. The construct used in this example (M1K1) is similar in design, but the tether and guide sequence have been changed to eliminate the possibility of contamination by previously evolved DNA-cleaving variants. The tethered 3′-guide sequence base-pairs with a complementary DNA oligonucleotide substrate to form a hybrid 16 bp helical stem that is recognized by M1 RNA [Forster and Altman, *Science* 249(4970):783-786 (1990); Frank et al., *Biochemistry* 33:10800-10808 (1994); Li and Altman, *Nucleic Acids Res.* 24(5):835-842 (1996); Kikuchi et al., *Nucleic Acids Res* 21(20):4685-9 (1993)]. A 3′-ACCA motif in the guide sequence directs the cleavage site [Guerrier-Takada et al., *Cell* 38(1):219-224 (1984); Kirsebom and Svard, *Nucleic Acids Res* 20(3):425-432 (1992)] which lies towards the 5′-end of the M1K1/substrate helix. Selection is performed by annealing a population of M1K1 variants to biotinylated DNA substrate which is then immobilized on paramagnetic beads (Figure 8). Variants that cleave the DNA substrate in the appropriate buffer are recovered in the supernatant and used to seed the next generation.

The initial population of M1K1 variants was created using six consecutive rounds of mutagenic PCR (adapted from Cadwell and Joyce, 1994) which resulted in a heterogeneous pool of molecules containing an average of 16.3 mutations/molecule with a range of 12-21

mutations/molecule. A starting population of 10^{13} M1K1 molecules was selected in the presence of C5 protein with a 1:1 stoichiometric ratio of M1K1 RNA:C5 protein, thus enabling every RNA to couple with a protein subunit to form the holoenzyme complex [Talbot and Altman, *Biochemistry* 33(6):1399-405 (1994)].

Five generations of selection were performed as illustrated in Figure 8 and the evolved population (G5) was then assayed for DNA cleavage. Single-turnover kinetics were used to asses catalytic activity because product release is likely to be the rate limiting step in this *in vitro* selection system [Robertson and Joyce, *Nature* 344(6265):467-468 (1990); Tallsjo and Kirsebom, *Nucleic Acids Res* 21(1):51-57 (1993); Herschlag, *Proc Natl Acad Sci USA* 88(16):6921-6925 (1991)], and single-turnover conditions should largely eliminate the product release component of the first-order rate constant (k_{cat}). Under these conditions the G5 holoenzyme population showed a approximately 50-fold improvement in DNA cleavage efficiency (k_{cat}/K_M) compared to the wild-type holoenzyme (M1K1 RNA+C5 protein).

Selection for DNA cleavage was continued and improvement in activity reached a plateau after eight generations (G8). Two rounds of mutagenic PCR were performed after G8 to introduce additional genetic variation and gain further improvements in DNA cleavage activity. Two additional rounds of selection were performed after this mutagenesis (for a total of 10 generations of selection) which yielded minor improvements over G8 cleavage activity.

A sample of individual ribozymes from population G10 were cloned and assayed for DNA cleavage. Of the assayed variants, clone G10.5 was chosen for further characterization based on two criteria. First, it displayed DNA cleavage activity comparable to the G10 population average. Second, clone G10.5 has the fewest mutations of any clone isolated from the G10 population (relative to the wild-type sequence), thus improving the chance of linking sequence changes to the observed catalytic enhancement. Clone G10.5 contains 5 mutations (C29U, A136U, U284C, A333U, A351G) compared to the population average of 8.7 mutations/molecule.

Kinetic analysis shows that G10.5 holoenzyme (G10.5 RNA combined with C5 protein) cleaves the target DNA substrate approximately 100 times more efficiently (k_{cat}/K_M) than does the wild-type holoenzyme (Table 6). This increased efficiency is the result of a large increase (> 400-fold) in the first-order rate constant (k_{cat}) which is offset by a modest reduction in substrate binding (5-fold increase in K_M). Since the single-turnover kinetics employed in these analyses eliminate the product-release component of the kinetic pathway, these results suggest the improvement in DNA cleavage is due to an acceleration of the hydrolysis of the phosphodiester bond.

Table 6: Kinetic parameters of DNA cleavage for M1K1 RNA and evolved clone G10.5.

	Holoenzyme		
	k_{cat} (min^{-1})	K_M (nM)	k_{cat}/K_M ($min^{-1}M^{-1}$)
M1K1 (wt)	$2.1\pm0.3\times10^{-4}$	1.6 ± 0.4	$1.3\pm0.2\times10^{6}$
G10.5	$8.8\pm0.8\times10^{-2}$	7.8 ± 2.1	$1.1\pm0.2\times10^{8}$
	RNA only		
	k_{cat} (min^{-1})	K_M (nM)	k_{cat}/K_M ($min^{-1}M^{-1}$)
M1K1 (wt)	$1.0\pm0.5\times10^{-6}$	52 ± 7	$1.9\pm0.6\times10^{2}$
G10.5	$1.4\pm0.3\times10^{-6}$	203 ± 16	$6.9\pm1.0\times10^{1}$

2. Analysis of Mutations that Enhance DNA Cleavage

The functional effects of the mutations that characterize clone G10.5 were investigated, emphasizing two mutations that were ubiquitous in the sequences from the generation 10 (G10) population: A136U and A333U. The DNA cleavage enhancing effect of these prevalent mutations was assessed using a point-mutation strategy. Candidate mutations were placed in a wild-type M1K1 background and also reverted to the wild-type

nucleotide in an evolved clone (G10.5) background. These analyses revealed that both A136U and A333U play a major role in accelerating DNA cleavage. Of the two prevalent mutations, A333U had the greatest effect on DNA cleavage. When U333 is placed in a wild-type M1K1 background, DNA cleavage efficiency (k_{cat}/K_M) of the holoenzyme improved 28-fold (Figure 9 and Table 7). Conversely, reversion of this mutation to the wild-type A333 in clone G10.5 caused a 10-fold decrease in DNA cleavage efficiency of the holoenzyme.

The second dominant mutation, A136U, also had a significant effect on DNA cleavage. Placing this mutation in the M1K1 background produced a 7-fold improvement in holoenzyme DNA cleavage activity, while reversion of this mutation in clone G10.5 to the wild-type A136 resulted in a 3-fold reduction in holoenzyme activity (Table 7 and Table 8). Finally, placing both mutations simultaneously in a M1K1 background resulted in a 47-fold improvement in holoenzyme activity, while reversion of both mutations in G10.5 yielded a dramatic 480-fold decrease in activity. Reverting both A136U and A333U in the evolved background reduced DNA cleavage activity below that of wild-type.

As Figure 9 makes clear, the effect of the point mutations depends on the background in which they are placed (or reverted). In some cases, the effects of the sequence changes are nearly additive: in the wild-type background, U136 provides a 7-fold improvement and U333 a 28-fold improvement, while jointly they result in a 47-fold improvement in DNA cleavage. Placing both A136U and A333U simultaneously in a wild-type background, however, does not confer DNA-cleavage activity equal to that of clone G10.5 (Table 7). One or both of the prevalent mutations interact with additional changes in the G10.5 RNA molecule, and specifically with a third mutation (A351G). The effect of the prevalent mutations is also not symmetrical: reversions at positions 136 and 333 in the evolved background result in a 480-fold drop in activity (Table 7 and Table 8). These data suggest that a third mutation may be playing a significant role in facilitating DNA cleavage. Furthermore, this mutation becomes detrimental to cleavage

activity in the absence of U136 and U333, as demonstrated by the lower than wild-type activity of the double mutant G10.5A136A333.

Table 7: Kinetic parameters of DNA cleavage for M1K1 RNA and point mutation constructs

	RNA alone		
	k_{cat}x10^{-5} (min^{-1})	K_M (nM)	k_{cat}/K_M x10^5 ($min^{-1}M^{-1}$)
M1K1 (wt)	0.1±0.05	52 ± 7	0.019±0.006
M1K1U136	4.5±0.8	173±14	0.26±0.05
M1K1U333	0.12±0.03	55±5	0.022±0.004
M1K1U136U333	n.d.	n.d.	0.28±0.02
G10.5	0.14±0.03	203±16	0.007±0.001
G10.5A136	n.d.	n.d.	n.d.
G10.5A333	n.d.	n.d.	n.d.
G10.5A136A333	n.d.	n.d.	n.d.

n.d.= not determined

	Holoenzyme		
	k_{cat} x10^{-5} (min^{-1})	K_M (nM)	k_{cat}/K_M x10^5 ($min^{-1}M^{-1}$)
M1K1 (wt)	21±3	1.6 ±0.4	13±2
M1K1U136	620±110	6.6±1.2	94±8
M1K1U333	540±200	1.5±0.4	360±90
M1K1U136U333	5800±1000	9.5±1.5	610±90
G10.5	8800±800	7.8±2.1	1100±200
G10.5A136	1900±300	5.8±1.1	330±20
G10.5A333	1200±300	10.5±1.8	120±20
G10.5A136A333	150±40	6.5±1.5	2.3±1.0

Table 8: Relative DNA cleavage efficiency of point mutation constructs

	RNA alone	Holoenzyme
M1K1 (wt)	1	1
M1K1U136	12	7
M1K1U333	1	28
M1K1U136U333	15	47
G10.5	0.3	85
G10.5A136	n.d.	25
G10.5A333	n.d.	9
G10.5A136A333	n.d.	0.18

Relative efficiency is determined from the ratio of specificity constants (k_{cat}/K_M) using M1K1 wild-type as the denominator.
n.d.= not determined

Table 9. ANOVA of the catalytic efficiency (k_{cat}/K_M) of point-mutation constructs

MAIN EFFECTS	F ratio	Probability>F
Position 136	5.13	0.05 *
Position 333	16.26	0.003 **
Presence/absence of C5	23.84	0.0009 **
INTERACTIONS		
136 * 333	2.33	0.162
Protein * 136	5.12	0.05*
Protein * 333	16.26	0.003**

* significant at P 0.05
** significant at P 0.01

5

3. Interactions with C5 Protein

C5 protein interacts with M1K1 RNA to significantly increase the rate of DNA cleavage under selection conditions (Table 6). This increase in

activity indicates that the tethered guide sequence of the construct does not significantly interfere with C5 binding or formation of the holoenzyme complex. The wild-type M1K1 holoenzyme shows a 200-fold increase in the first-order rate constant k_{cat} for DNA cleavage and a 33-fold improvement in substrate binding (reduced K_M) compared to the M1K1 RNA-only reaction (Table 6). Overall, the addition of C5 results in a 6000–fold increase in DNA cleavage efficiency (k_{cat}/K_M) for the M1K1 holoenzyme.

In this example the RNase P holoenzyme was selected for improved DNA cleavage. The design of the method allows the RNA moiety to evolve while maintaining the sequence of the C5 protein. Under these selection conditions, the ribozyme could evolve a set of mutations that enhance DNA cleavage independent of the presence or absence of the protein component. Alternatively, the three-dimensional configuration adopted by the DNA-cleaving holoenzyme derivatives may depend entirely on the continued interaction with the C5 component.

The character of the RNA-protein interaction in the evolved molecules was explored by assaying clone G10.5 and several point mutation constructs in the presence and in the absence of C5 protein. Clone G10.5 shows a dramatic enhancement in DNA cleavage when complexed with C5 protein: the holoenzyme shows a 6300-fold increase in k_{cat} and a 26-fold improvement in substrate binding (decreased K_M) compared to the G10.5 RNA-only reaction (See Figure 4 and Table 1). Thus, the presence of C5 protein has a greatest effect on the evolved clone: G10.5 shows a 10^6–fold increase in DNA cleavage efficiency (k_{cat}/K_M) when complexed with C5, compared to a 6000-fold increase for wild-type M1K1.

The dramatic acceleration in DNA cleavage produced by C5 protein can be dissected by careful examination of the point mutation constructs. Kinetic analysis reveals that the beneficial effect of mutation A333U is only apparent in the presence of C5 protein (Tables 7 and 8). In contrast, the second prevalent mutation, A136U, enhances DNA cleavage both in the presence and in the absence of C5 protein (Table 8).

Since the increase in DNA-cleavage produced by mutation A333U depends on the presence of the C5 protein, the possibility that this mutation increases M1K1 RNA-C5 protein affinity was assessed. If U333 increases RNA-protein affinity, it might act by enhancing the formation of the holoenzyme. The dissociation constant (K_d) for C5 protein and wild-type M1K1 RNA, as well as for several evolved RNAs, were determined using a gel-retardation method (Talbot and Altman, 1994). These analyses indicated that the affinity between C5 protein and evolved RNAs, specifically G10.5, is not significantly different from the affinity between C5 and wild-type M1K1: the K_d for wild-type M1K1 is 0.5±0.1 nM while the K_d for evolved clone G10.5 is 0.7±0.2 nM.

Interactions between the fixed sequence changes, the ribozyme background and C5 protein were probed via an analysis of variance (ANOVA) of construct DNA cleavage efficiencies (k_{cat}/K_M). In this analysis, the relative contribution of the individual variables (e.g., the individual point mutations, the ribozyme background and the presence or absence of the protein) can be quantified as "main effects". In addition, this technique also explores the joint effect of two or more variables ("interaction effects") on the full set of cleavage assays.

The ANOVA results are shown in Table 9. As expected, the identities of the nucleotides at position 136 (p=0.05) and position 333 (p=0.003) significantly affect the catalytic efficiency of both the RNA-only and holoenzyme constructs. This analysis also underscores the profound effect (p=0.0009) of the presence/absence of the C5 protein on the performance of evolved and reverted ribozyme constructs. This approach reveals a very strong interaction between position 333 and the C5 protein (p=0.003) and a marginally significant interplay between position 136 and the C5 protein (p=0.05). No significant interaction is apparent between positions 136 and 333.

C. Discussion

In vivo, RNase P substrate versatility is affected by the C5 protein in several respects. First, the holoenzyme cleaves pre-tRNA, pre-4.5S RNA

and other RNA substrates with similar efficiencies, whereas M1 RNA cleaves pre-tRNA most efficiently [Altman et al., *Faseb J* 7(1):7-14 (1993); Hartmann et al., *Proc Natl Acad Sci USA* 92:5822-5826 (1995); Peck-Miller and Altman, *J Mol Biol* 221(1):1-5 (1991); Niranjanakumari et al., *Proceedings of the National Academy of Sciences of the United States of America* 95(26):15212-15217 (1998)]. The activity-normalizing effect of the C5 protein results primarily from an increased affinity for precursor RNA substrates without an accompanying increase in product affinity [Crary et al., *Biochemistry* 37(26):9409-16 (1998); Kurz et al., *Biochemistry* 37(8):2393-2400 (1998)]. For example, the holoenzyme increases affinity for pre-tRNA 10^4-fold and pre-4.5S RNA approximately 10^2-fold without making product release overly rate-limiting [Fierke et al., *FASEB Journal* 13(7) (1999); Peck-Miller and Altman, *J Mol Biol* 221(1):1-5 (1991)]. Second, the presence of C5 protein alters interactions with the pre-tRNA 5′-leader sequence and T stem [Crary et al., *Biochemistry* 37(26):9409-16 (1998); Niranjanakumari et al., *Proceedings of the National Academy of Sciences of the United States of America* 95(26):15212-15217 (1998)]. Third, the protein component lowers the concentration of magnesium required for efficient catalysis [Guerrier-Takada et al., *Cell* 35(3 Pt 2):849-57 (1983); Reich et al. *Science* 239(4836):178-181 (1988)]. Finally, the C5 protein subunit also appears to mitigate structural defects in M1 RNA. Certain M1 RNA mutants defective in ptRNA cleavage, for example, regained the ability to cleave ptRNA in the presence of the protein [Lumelsky and Altman, *J Mol Biol* 202(3):443-54 (1988)].

1. RNase P Evolves Efficient DNA Cleavage

In this example, modified version of RNase P from *E. coli* was selected for cleavage of a novel DNA substrate. The RNase P holoenzyme was subjected to *in vitro* evolution, but allowed only the catalytic RNA subunit of the enzyme to evolve (the C5 protein component remained constant). *In vitro* selection for DNA cleavage was accomplished by using a construct of M1 RNA (M1K1) containing a 3'-tethered guide sequence that is complementary to the DNA target. A population of RNase P holoenzyme

variants responded rapidly to *in vitro* selection and after ten generations (G10) showed approximately 100-fold enhancement in DNA cleavage efficiency (k_{cat}/K_M) compared to the wild-type M1K1 holoenzyme. This cleavage enhancement results in a highly proficient DNA-cleaving enzyme; in fact, the evolved clone G10.5 holoenzyme cleaves DNA almost as efficiently as wild-type RNase P cleaves ptRNA, under selection conditions (Tables 1 and 5). With a DNA cleavage specificity constant on the order of 10^8, the G10.5 holoenzyme compares favorably with many protein catalysts.

A detailed kinetic analysis of evolved clone G10.5 reveals an increase in the first-order rate constant (k_{ca}) of DNA cleavage which is offset by a reduction in substrate affinity (K_M) (Table 6). Because single-turnover kinetics minimize the product-release component of k_{ca}, this enhancement results from an acceleration of the chemistry of cleavage, rather than increased product release [Cole and Dorit, *Journal of Molecular Biology* 292(4):931-944 (1999)].

Sequences from the final generation of *in vitro* evolution showed two ubiquitous mutations, A136U and A333U, that can be directly linked to enhanced DNA cleavage. These mutations were mapped onto a provisional model of M1K1 RNA. This mapping places the mutations in two separate structural domains of the ribozyme. Mutation A136U is found in the single stranded region J11/12 which connects helices P11 and P12 in the substrate recognition domain (Domain I, Loria and Pan,1996)[Loria and Pan, *RNA* 2(6):551-563 (1996)]. Point mutation analysis indicates that A136U increases the rate of DNA cleavage while simultaneously reducing the rate of ptRNA cleavage (Table 9). The second ubiquitous mutation, A333U, occurs within the catalytic domain of M1 RNA, in region J18/2 (Domain II, Loria and Pan,1996).

Whereas mutation A136U enhances DNA cleavage in the presence and absence of C5 protein, A333U requires C5 protein to enhance DNA cleavage: when placed in a wild-type M1K1 background mutation A333U neither helps, nor hinders, DNA cleavage in the absence of C5 protein (Tables 7 and 8). Upon addition of C5 protein however, the mutation confers

a considerable improvement in DNA cleavage. This dependence on C5 protein for activity suggests that either A333U directly interacts with C5 protein or that the protein causes a conformational change in M1 RNA that enables A333U to enhance DNA cleavage.

In summary, improvements in DNA cleavage can be traced primarily to two distinct mutations. Mutation A136U lies in the substrate recognition domain of M1K1 RNA, affects DNA cleavage independently of C5 protein and appears to be mechanistically similar to mutations that evolve in the absence of C5 protein. Mutation A333U lies in close proximity to the catalytic core of the ribozyme, enhances DNA cleavage only in the presence of C5 protein and accelerates the rate of cleavage without affecting substrate binding.

2. The Role of C5 Protein in Substrate Versatility

The approximately 100-fold improvement in DNA cleavage efficiency (k_{cat}/K_M) exhibited by RNase P evolved *in vitro* emphasizes the inherent substrate versatility of this enzyme. Although C5 protein is the cornerstone of RNase P substrate versatility *in vivo*, it did not mitigate the loss of versatility associated with evolution *in vitro*. The tradeoff pattern generated by evolving the holoenzyme is very similar to that obtained when evolving M1 RNA in the absence of the protein cofactor. Specifically, evolved clones maintain catalytic activity on p4.5S RNA and on an RNA oligo of identical sequence (both long helical substrates similar in shape to the DNA substrate), while losing activity on the ptRNA substrate. Here this tradeoff can be traced directly to mutation A136U: this mutation improves DNA cleavage while crippling ptRNA cleavage. Reversion of this mutation in clone G10.5 restores wild-type levels of ptRNA cleavage in the presence and absence of C5 protein. Mutations producing a similar tradeoff occur in M1 RNA evolved for DNA cleavage in the absence of C5 protein. Although C5 is important for substrate versatility, the protein cannot buffer the consequences of specialization that occur when evolving M1 RNA for DNA cleavage.

The tradeoff between DNA cleavage and ptRNA cleavage generated in the selection may be caused by the phenomenon of phenotypic drift. Since mutations that improve a single enzymatic trait at the expense of another trait are common, and mutations that improve more than one trait are rare, any trait which is not under selection will decay due to drift [Giver et al., *Proceedings of the National Academy of Sciences of the United States of America* 95(22):12809-13 (1998)]. Thus, it should be possible to evolve DNA cleavage without a tradeoff in versatility by simultaneously selecting for cleavage of both DNA and ptRNA. This dual-activity selection regime would necessarily eliminate mutations such as U136 in the J11/12 region that interact with ptRNA.

Example 9: Molecular Modeling of M1K1

In an effort to understand how mutations A136U and A333U affect DNA cleavage, these changes were positioned on a 3-D structural model of M1K1 RNA. This model was constructed using the program MANIP, designed by Massire and Westhof (1998)[Massire and Westhof, *Journal of Molecular Graphics and Modelling* 16(4-6):197-205; 255-7 (1998)], which allows the manipulation and positioning of RNA structures.

The PDB file of the Massire et al., *J. Mol. Biol.* 279:773-793 (1998), model of M1 RNA was obtained from the RNase P data base (URL: http://www.mbio.ncsu.edu/RNaseP/home.html). The program MANIP (Massire and Westhof 1998) was used to append an RNA tether to the 3' terminus of the M1 RNA model (tether and guide sequence: 5' –M1 RNA-ACCUAAUUUACCCCAUCCAUGACACUCCUGCGCAGACACGGCCU CACCAG-3', SEQ ID NO:30). An A-form heteroduplex between the RNA guide sequence and DNA substrate was made (DNA: 5'-TCCTACCTGCTAACTGAGGCCGTGTCTGCG-3', SEQ ID NO:31). This RNA/DNA helix was then superimposed on the acceptor stem of the $ptRNA^{Tyr}$ substrate such that the helices and cleavage sites were aligned. The M1K1 model was then checked for inappropriate contacts between the substrate helix and the ribozyme.

The model of the DNA oligo substrate docked in the catalytic cleft of M1K1 RNA was constructed by adding a 3′-tethered RNA guide sequence to the existing model of M1 RNA by Massire et al., (1998). The RNA guide sequence/DNA oligo helix was created and refined using MANIP and linked to the 3'-end of M1 RNA. Positioning the A-form heteroduplex was accomplished by superimposing the helix on the ptRNA acceptor stem of the existing model [Massire et al., *J. Mol. Biol.* 279:773-793 (1998)]. This docking strategy allowed alignment of the ptRNA and DNA cleavage sites without creating any inappropriate contacts between the ribozyme and substrate.

Mutation A136U lies in the single-stranded region J11/12 of M1 RNA and corresponds to a universally conserved position (A) in known RNase P RNA sequences from Eubacteria [Brown, *NAR* 24:236 (1996); Haas and Brown, *Nucleic Acids Research* 26(18):4093-4099 (1998)]. The M1K1 RNA model places this mutation in very close proximity (less than 8 angstroms) to the 3′-terminus of our DNA oligo substrate. The J11/12 region of the ribozyme is know to be important for ptRNA binding and has been shown to be in close proximity to the conserved –GTΨCR- motif of the ptRNA T loop [Harris et al., *RNA* **3**(6):561-576 (1997); Loria and Pan, *Biochemistry* 36(21):6317-6325 (1997); Kirsebom and Vioque, *Molecular Biology Reports* 22:99-109 (1996)].

Mutation A333U, located in the single-stranded region J18/2, also occurs in a universally conserved position in all eubacterial RNase P RNA sequences. The M1K1 RNA model places this nucleotide near the 5′-terminus of the DNA oligo substrate very close to the site of cleavage. Region J18/2 is know to be important for catalytic activity and has been show to lie in close proximity to the 5'-leader of ptRNA [Harris et al., *EMBO* 13(17):3953-3963 (1994); Harris and Pace, *Mol Biol Rep* 22(2-3):115-123 (1995)].

CLAIMS

We claim:

1. A nucleic acid molecule comprising

a variant RNase P RNA, wherein the variant RNase P RNA cleaves a DNA substrate with a catalytic efficiency greater than wild-type RNase P RNA.

2. The nucleic acid molecule of claim 1 wherein the variant RNase P RNA has one or more of the following alterations: a C at position 18, an A at position 19, a U at position 26, a C at position 28, a U at position 59, an A at position 87, an A at position 101, a U at position 129, an A at position 137, a U at position 138, a U at position 139, a G at position 173, an A at position 203, a U at position 226, a U at position 228, a C at position 270, a C at position 299, a G at position 337, a C at position 362, a G at position 371, a G at position 386, any base at position 390, and an A at position 396.

3. The nucleic acid molecule of claim 1 wherein the variant RNase P RNA contains one or more of the following mutations: a C at position 18, a U at position 26, a U at position 59, a U at position 129, an U at position 138, a U at position 139, a G at position 173, an A at position 203, a U at position 228, a C at position 270, a C at position 299, a G at position 371, and an A at position 396.

4. A set of nucleic acid molecules comprising

a plurality of nucleic acid molecules according to claim 1.

5. The set of nucleic acid molecules of claim 4 wherein the nucleic acid molecules of the set cleave DNA substrates with a range of catalytic efficiencies greater than wild-type RNase P RNA.

6. A vector encoding the nucleic acid molecule of claim 1.

7. A cell comprising the vector of claim 6.

8. A cell comprising the nucleic acid molecule of claim 1.

9. A set of nucleic acid molecules comprising

a plurality of variant RNase P RNA, wherein the variant RNase P RNA cleave a DNA substrate with a catalytic efficiency greater than wild-type RNase P RNA.

10. A nucleic acid molecule isolated by

(a) generating a population of nucleic acid molecules comprising variant nucleic acid molecules, wherein each variant nucleic acid molecule comprises a variant RNase P RNA molecule, wherein the sequence of the variant RNase P RNA molecule differs from the sequence of wild-type RNase P RNA;

(b) bringing into contact

(i) the population of nucleic acid molecules,

(ii) substrate sequences of interest, wherein the substrate sequences each comprise deoxyribonucleotides,

(iii) guide sequences, wherein the guide sequences hybridize to the substrate sequences of interest, and wherein the guide sequences can mediate cleavage of an RNA molecule by wild-type RNase P RNA, wherein the RNA molecule is composed of ribonucleotides and wherein the sequence of the RNA molecule corresponds to the sequence of the substrate sequences, and;

(c) separating nucleic acid molecules that cleave the substrate sequence from nucleic acid molecules that do not;

wherein the nucleic acid molecules that cleave the substrate sequence each comprise a variant RNase P RNA that cleaves a DNA substrate with a catalytic efficiency greater than wild-type RNase P RNA.

11. The nucleic acid molecule of claim 10 wherein each nucleic acid molecule is coupled to a guide sequence.

12. The nucleic acid molecule of claim 10 wherein the substrate sequence consists of deoxyribonucleotides.

13. The nucleic acid molecule of claim 10 wherein a capture tag is coupled to the substrate sequence.

14. The method of claim 10 wherein an RNase P protein subunit is also brought into contact with the population of nucleic acid molecules, the substrate sequences of interest, and the guide sequences in step (b).

15. The method of claim 14 wherein the RNase P protein subunit is the C5 protein.

16. A method of producing variant RNase P RNA that cleaves a DNA substrate with a catalytic efficiency greater than wild-type RNase P RNA, the method comprising

(a) generating a population of nucleic acid molecules comprising variant nucleic acid molecules, wherein each variant nucleic acid molecule comprises a variant RNase P RNA molecule, wherein the sequence of the variant RNase P RNA molecule differs from the sequence of wild-type RNase P RNA;

(b) bringing into contact

(i) the population of nucleic acid molecules,

(ii) substrate sequences of interest, wherein the substrate sequences each comprise deoxyribonucleotides,

(iii) guide sequences, wherein the guide sequences hybridize to the substrate sequences of interest, and wherein the guide sequences can mediate cleavage of an RNA molecule by wild-type RNase P RNA, wherein the RNA molecule is composed of ribonucleotides and wherein the sequence of the RNA molecule corresponds to the sequence of the substrate sequences, and;

(c) separating nucleic acid molecules that cleave the substrate sequence from nucleic acid molecules that do not;

wherein the nucleic acid molecules that cleave the substrate sequence each comprise a variant RNase P RNA that cleaves a DNA substrate with a catalytic efficiency greater than wild-type RNase P RNA.

17. The method of claim 16 wherein each nucleic acid molecule is coupled to a guide sequence.

18. The method of claim 16 wherein the substrate sequence consists of deoxyribonucleotides.

19. The method of claim 16 wherein a capture tag is coupled to the substrate sequence.

20. The method of claim 16 wherein nucleic acid molecules that cleave the substrate sequence are separated from nucleic acid molecules that do not by

associating the capture tag to a capture tag receptor, wherein the capture tag receptor is attached to a substrate, wherein nucleic acid molecules that do not cleave the substrate sequence become or remain associated with the substrate, wherein nucleic acid molecules that cleave the substrate sequence do not remain associated with the substrate, and

separating nucleic acid molecules not associated with the substrate from the nucleic acid molecules associated with the substrate.

21. The method of claim 16 wherein an RNase P protein subunit is also brought into contact with the population of nucleic acid molecules, the substrate sequences of interest, and the guide sequences in step (b).

22. The method of claim 21 wherein the RNase P protein subunit is the C5 protein.

23. A variant RNase P RNA, wherein the variant RNase P RNA cleaves a DNA substrate with a catalytic efficiency greater than wild-type RNase P RNA.

24. The variant RNase P RNA of claim 23 wherein the variant RNase P RNA has one or more of the following alterations: a C at position 18, an A at position 19, a U at position 26, a C at position 28, a U at position 59, an A at position 87, an A at position 101, a U at position 129, an A at position 137, a U at position 138, a U at position 139, a G at position 173, an A at position 203, a U at position 226, a U at position 228, a C at position 270, a C at position 299, a G at position 337, a C at position 362, a G at position 371, a G at position 386, any base at position 390, and an A at position 396.

25. The variant RNase P RNA of claim 23 wherein the variant RNase P RNA contains one or more of the following mutations: a C at position 18, a U at position 26, a U at position 59, a U at position 129, an U at position 138, a U at position 139, a G at position 173, an A at position 203, a U at position 228, a C at position 270, a C at position 299, a G at position 371, and an A at position 396.

26. A vector encoding the variant RNase P RNA of claim 23.

27. A cell comprising the vector of claim 26.

28. A cell comprising the variant RNase P RNA of claim 23.

29. A variant RNase P RNA isolated by

(a) generating a population of variant RNase P RNAs, wherein the sequence of each variant RNase P RNA differs from the sequence of wild-type RNase P RNA;

(b) bringing into contact

(i) the population of variant RNase P RNAs,

(ii) substrate sequences of interest, wherein the substrate sequences each comprise deoxyribonucleotides,

(iii) guide sequences, wherein the guide sequences hybridize to the substrate sequences of interest, and wherein the guide sequences can mediate cleavage of an RNA molecule by wild-type RNase P RNA, wherein the RNA molecule is composed of ribonucleotides and wherein the sequence of the RNA molecule corresponds to the sequence of the substrate sequences, and;

(c) separating variant RNase P RNAs that cleave the substrate sequence from variant RNase P RNAs that do not;

wherein the variant RNase P RNAs that cleave the substrate sequence are variant RNase P RNAs that cleaves a DNA substrate with a catalytic efficiency greater than wild-type RNase P RNA.

30. The method of claim 29 wherein each nucleic acid molecule is coupled to a guide sequence.

31. The method of claim 29 wherein the substrate sequence consists of deoxyribonucleotides.

32. The method of claim 29 wherein a capture tag is coupled to the substrate sequence.

33. The method of claim 29 wherein an RNase P protein subunit is also brought into contact with the population of variant RNase P RNAs, the substrate sequences of interest, and the guide sequences in step (b).

34. The method of claim 33 wherein the RNase P protein subunit is the C5 protein.

35. A method of cleaving a DNA substrate, the method comprising bringing into contact

(a) a DNA substrate of interest,

(b) a variant RNase P RNA, wherein the variant RNase P RNA cleaves DNA substrates with a catalytic efficiency greater than wild-type RNase P RNA, and

(c) a guide sequence, wherein the guide sequence hybridizes to the DNA substrate of interest.

36. The method of claim 35 wherein the variant RNase P RNA is coupled to the guide sequence.

37. The method of claim 36 wherein the DNA substrate is in a cell, wherein the variant RNase P RNA and the guide sequence are brought into contact with the DNA substrate by introducing the variant RNase P RNA and the guide sequence into the cell.

38. The method of claim 37 wherein the variant RNase P RNA and the guide sequence are introduced into the cell by introducing a vector encoding the variant RNase P RNA and the guide sequence into the cell.

39. The method of claim 38 wherein the vector expresses the variant RNase P RNA and the guide sequence as a single transcript.

40. The method of claim 35 wherein an RNase P protein subunit is also brought into contact with the DNA substrate of interest, the variant RNase P RNA, and the guide sequence.

41. The method of claim 40 wherein the RNase P protein subunit is the C5 protein.

42. A variant RNase P RNA isolated by

(a) generating a population of variant RNase P RNAs, wherein the sequence of each variant RNase P RNA differs from the sequence of wild-type RNase P RNA;

(b) bringing into contact

(i) the population of variant RNase P RNAs,

(ii) a substrate sequence of interest, wherein the substrate sequence comprises deoxyribonucleotides, and;

(c) separating variant RNase P RNAs that cleave the substrate sequence from variant RNase P RNAs that do not.

A

FIG. 1A

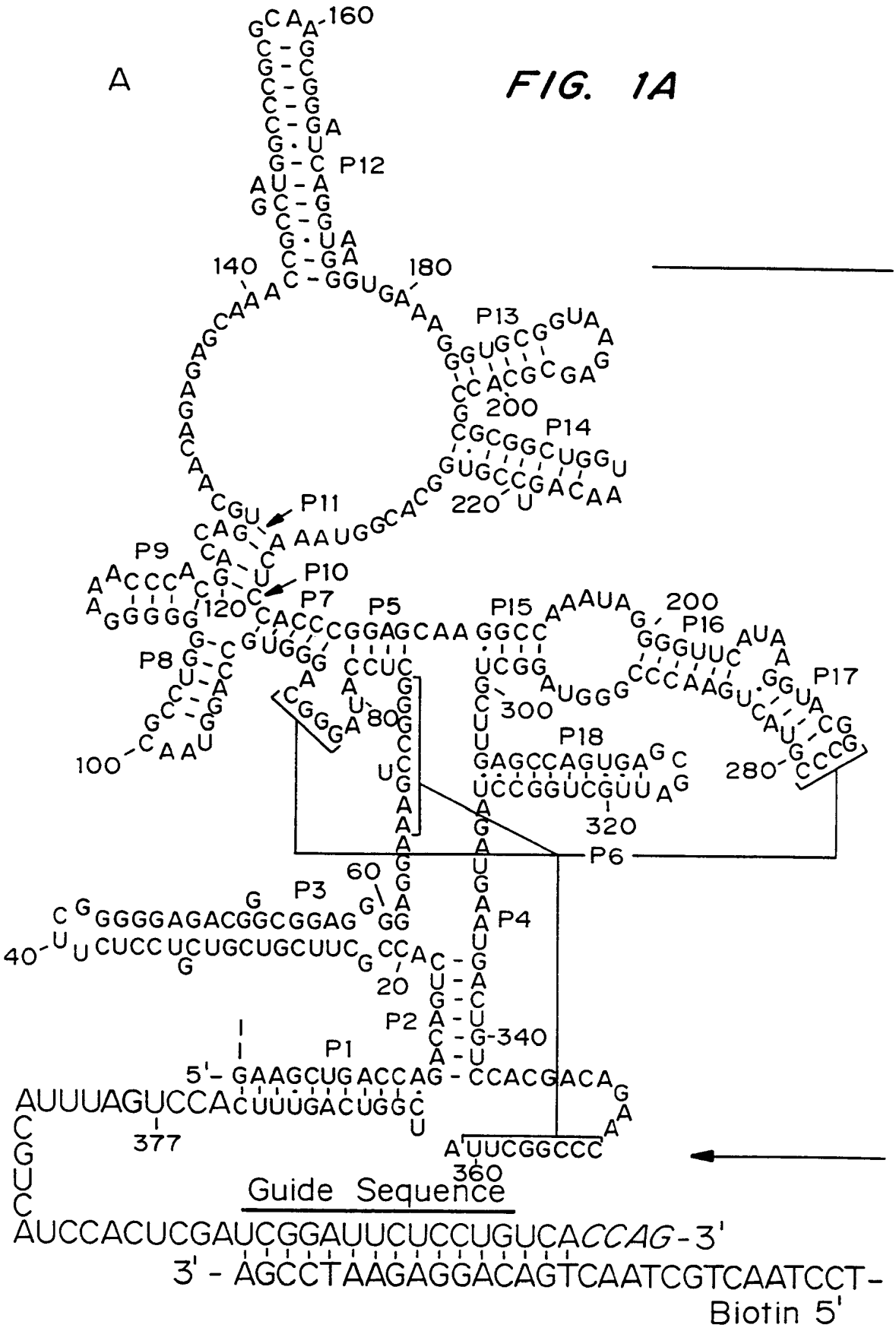
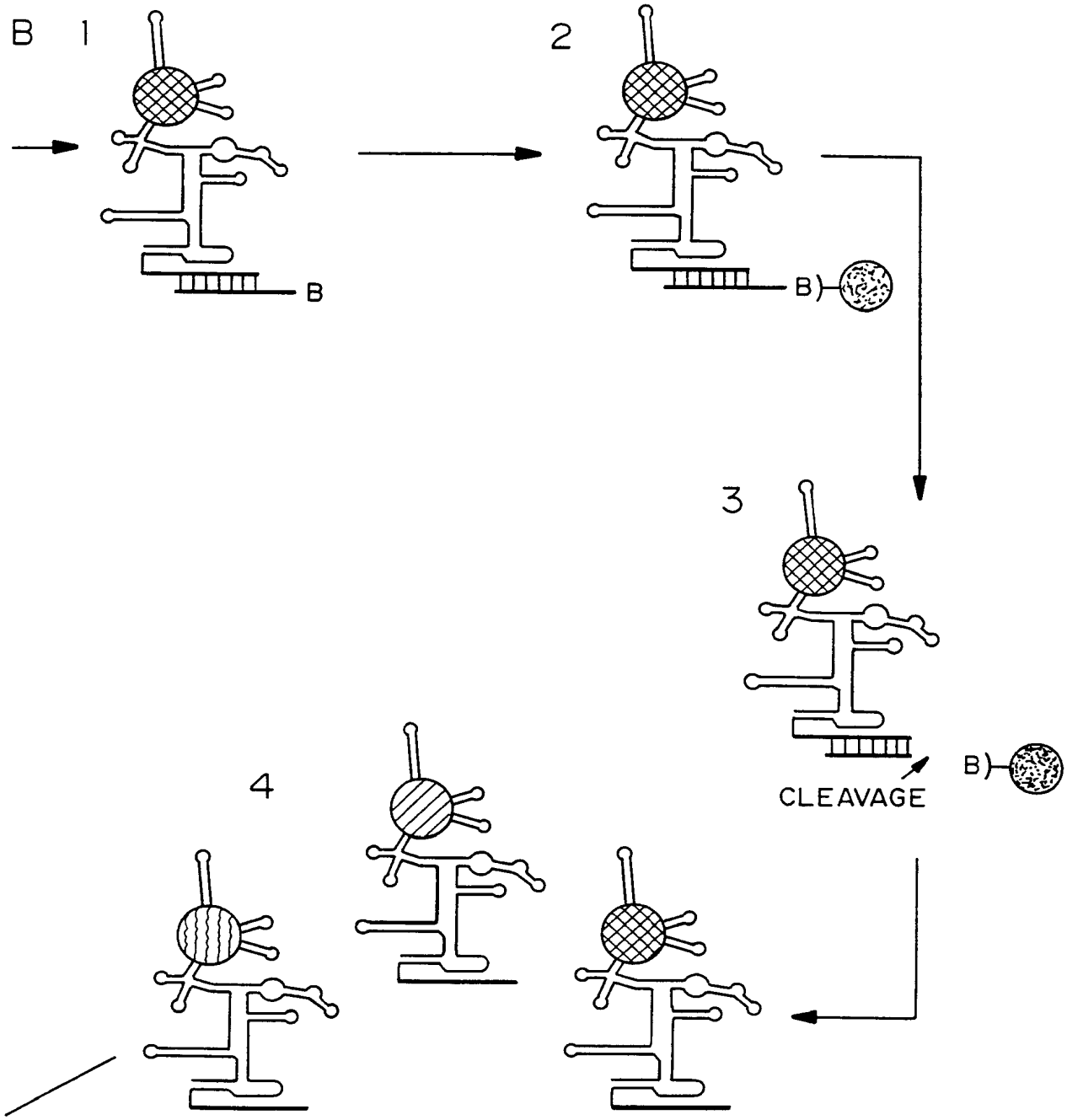

160
140
180
200
220
P1
P2
P3
P4
P5
P6
P7
P8
P9
P10
P11
P12
P13
P14
P15
P16
P17
P18
100
120
280
300
320
340
360
377
5'- GAAGCUGACCAG - CCACGACA
AUUUAGUCCACUUUGACUGGC
AUCCACUCGAUCGGAUUCUCCUGUCA
Guide Sequence
AUCCACUCGAUCGGAUUCUCCUGUCACCAG-3'
3' - AGCCTAAGAGGACAGTCAATCGTCAATCCT-
Biotin 5'

FIG. 1B



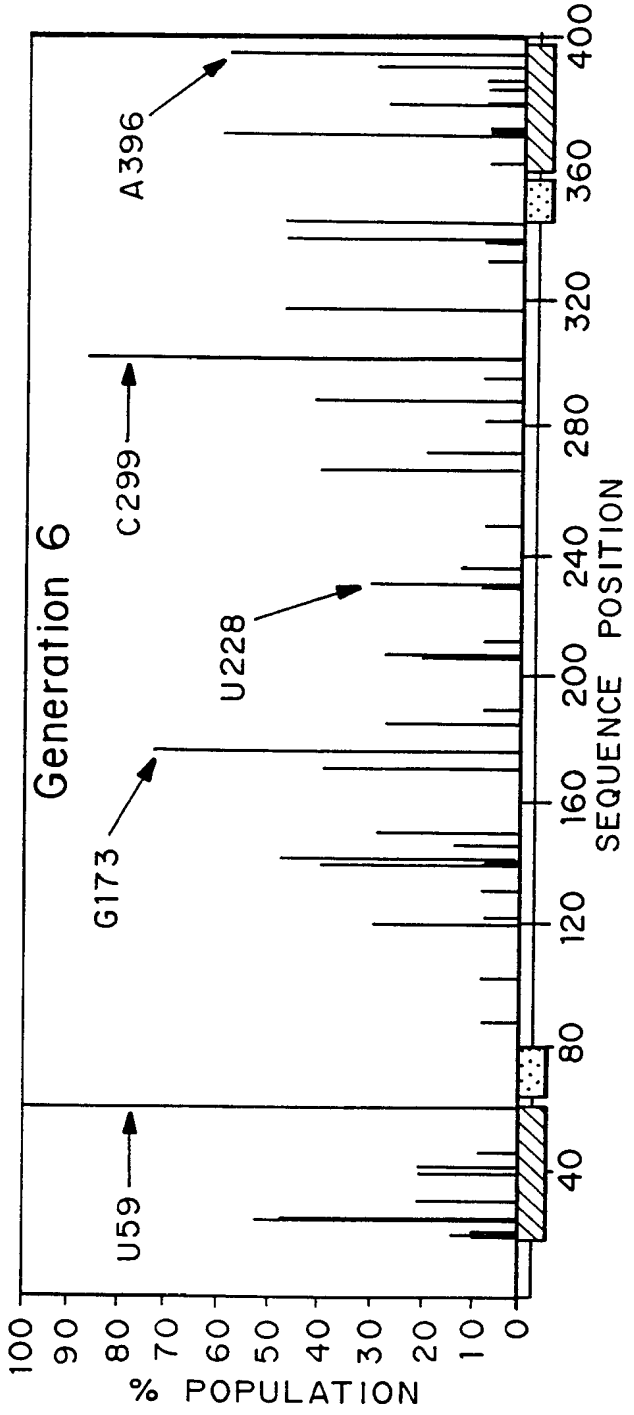

Generation 6
A396
C299
U228
G173
U59
% POPULATION
SEQUENCE POSITION
0
10
20
30
40
50
60
70
80
90
100
40
80
120
160
200
240
280
320
360
400

FIG. 2A

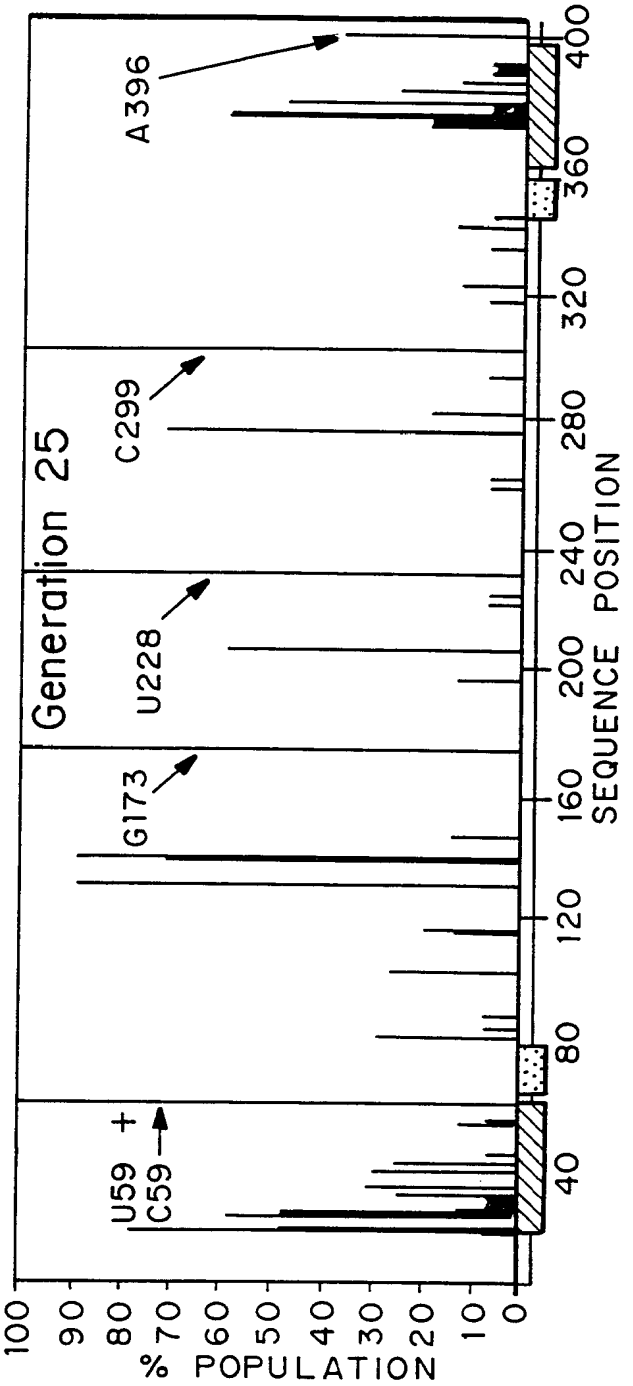

Generation 25
A396
C299
U228
G173
U59 +
C59
% POPULATION
SEQUENCE POSITION
0
10
20
30
40
50
60
70
80
90
100
40
80
120
160
200
240
280
320
360
400

FIG. 2B

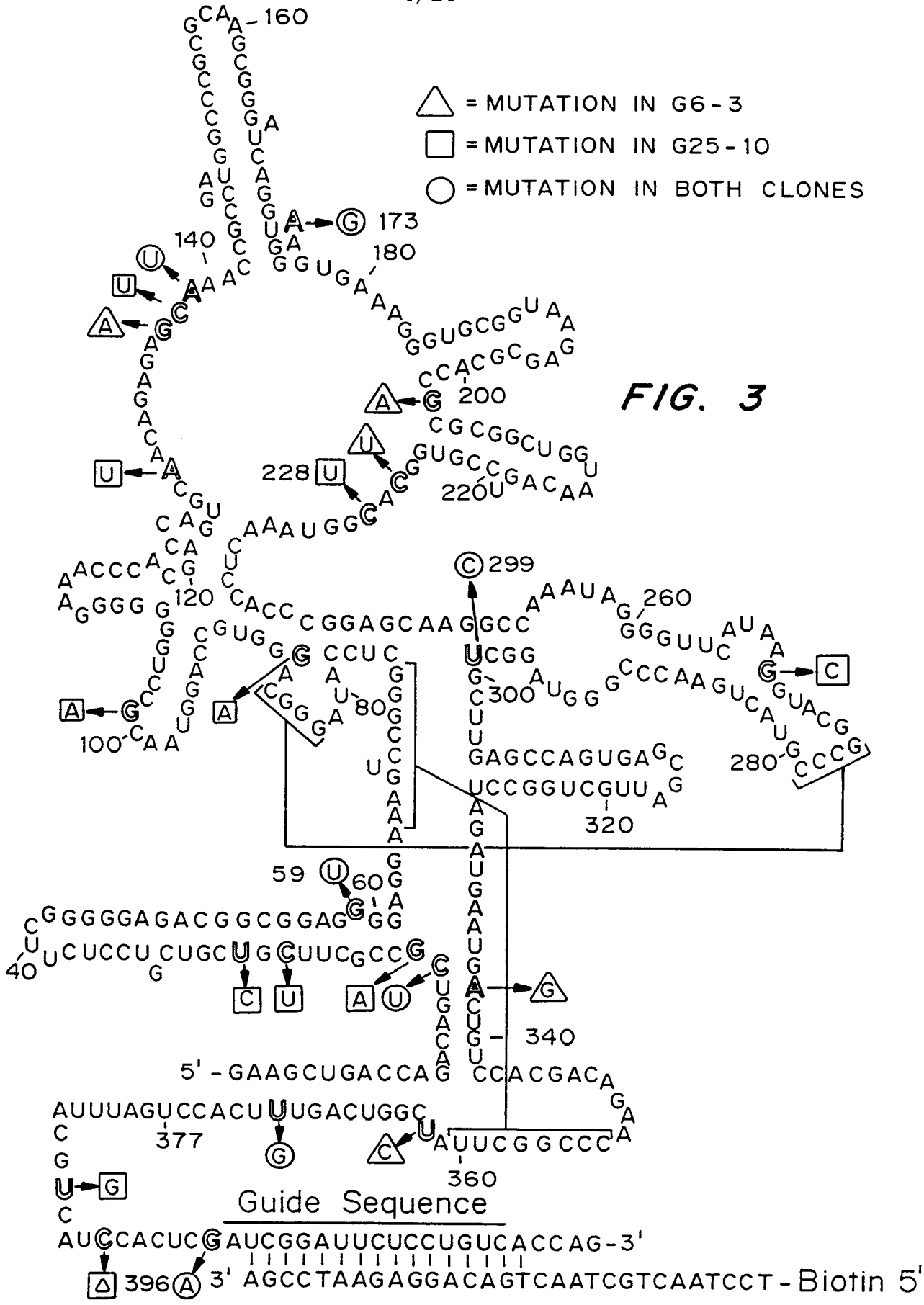

= MUTATION IN G6-3
= MUTATION IN G25-10
= MUTATION IN BOTH CLONES
Guide Sequence
5'-GAAGCUGACCAG CCACGACA
AUCCACUCGAUCGGAUUCUCCUGUCACCAG-3'
3' AGCCTAAGAGGACAGTCAATCGTCAATCCT-Biotin 5'

FIG. 3

FIG. 4

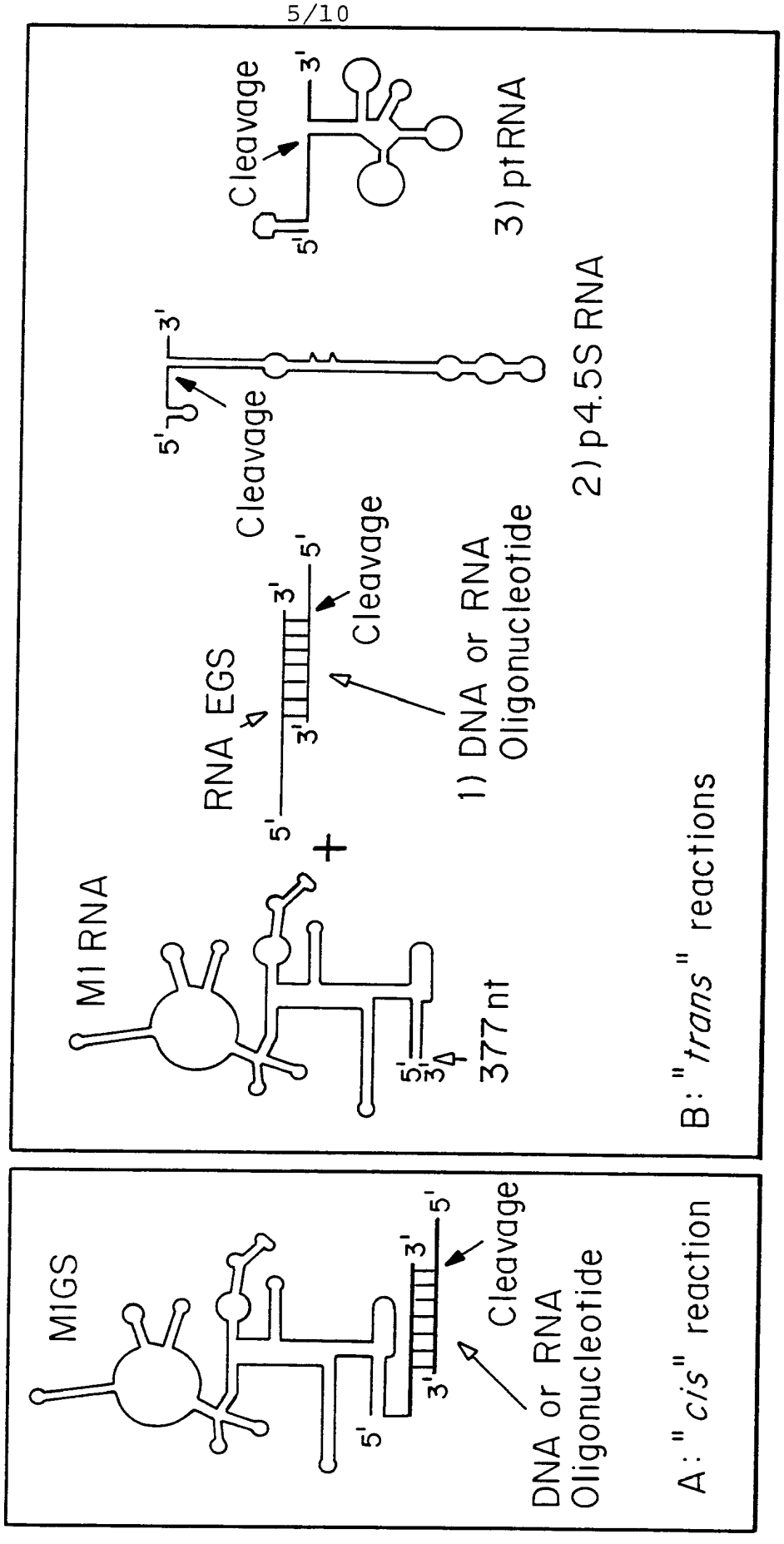

M1GS
5'
3'
3'
5'
Cleavage
DNA or RNA
Oligonucleotide
A: "cis" reaction
M1 RNA
377 nt
5'
3'
RNA EGS
5'
3'
3'
5'
Cleavage
1) DNA or RNA
Oligonucleotide
5'
3'
Cleavage
2) p4.5S RNA
Cleavage
5'
3'
3) ptRNA
B: "trans" reactions

FIG. 5

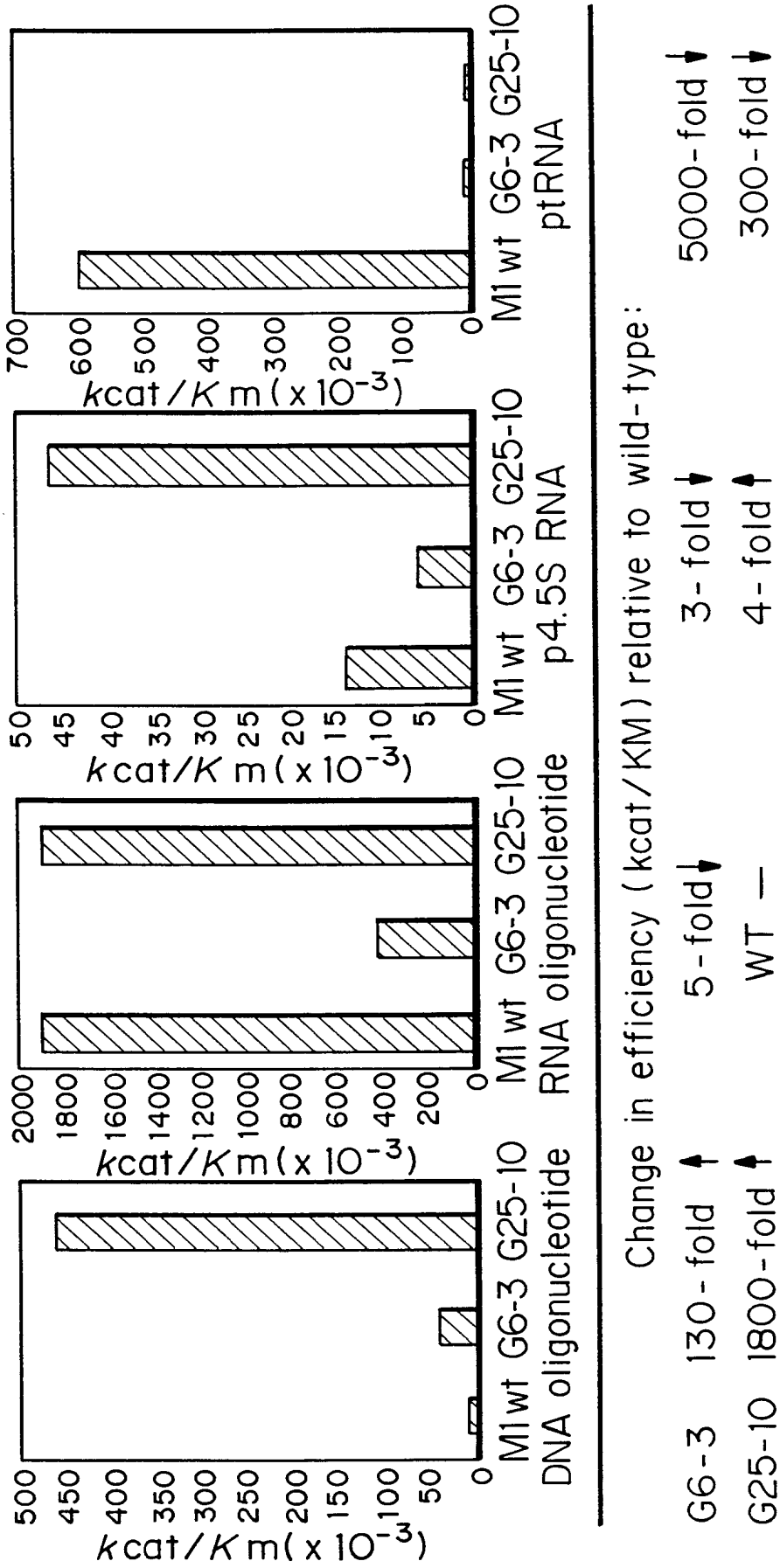

kcat/Km (x 10⁻³)
M1 wt G6-3 G25-10
DNA oligonucleotide
RNA oligonucleotide
p4.5S RNA
ptRNA
Change in efficiency (kcat/KM) relative to wild-type:
G6-3 130-fold 5-fold 3-fold 5000-fold
G25-10 1800-fold WT — 4-fold 300-fold

FIGURE 6

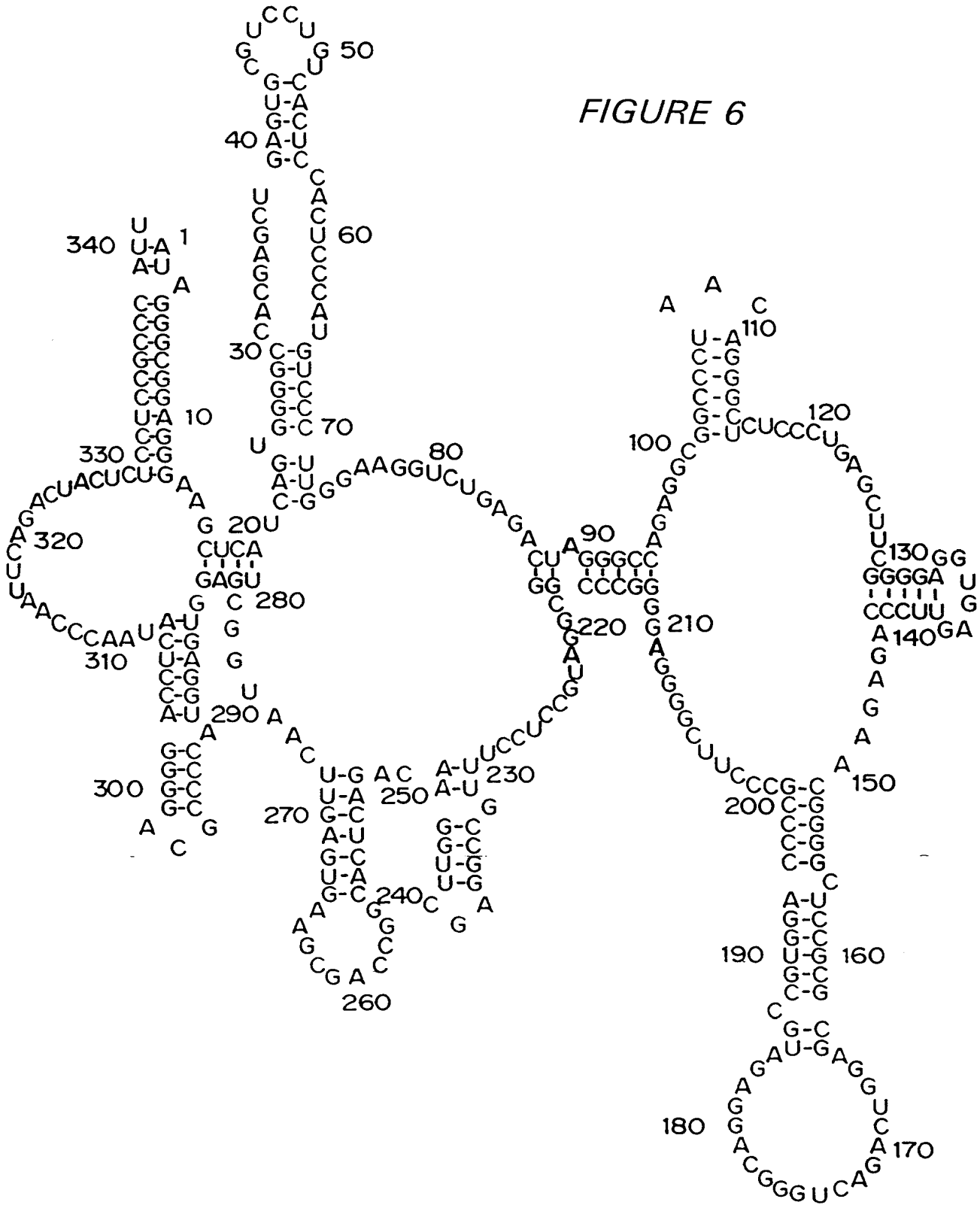
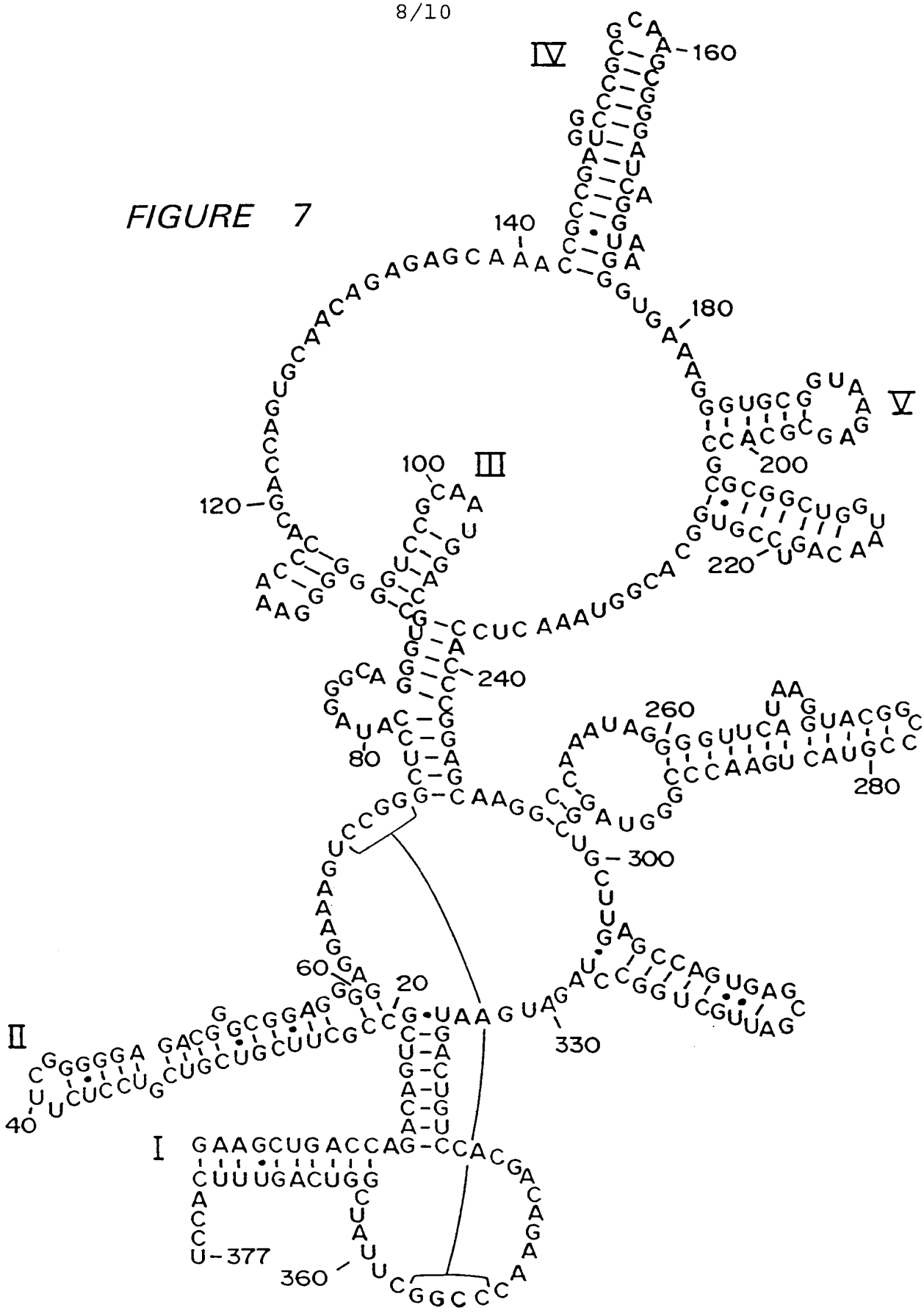


FIGURE 7


I
II
III
IV
V
40
60
80
100
120
140
160
180
200
220
240
260
280
300
330
360
20
U-377

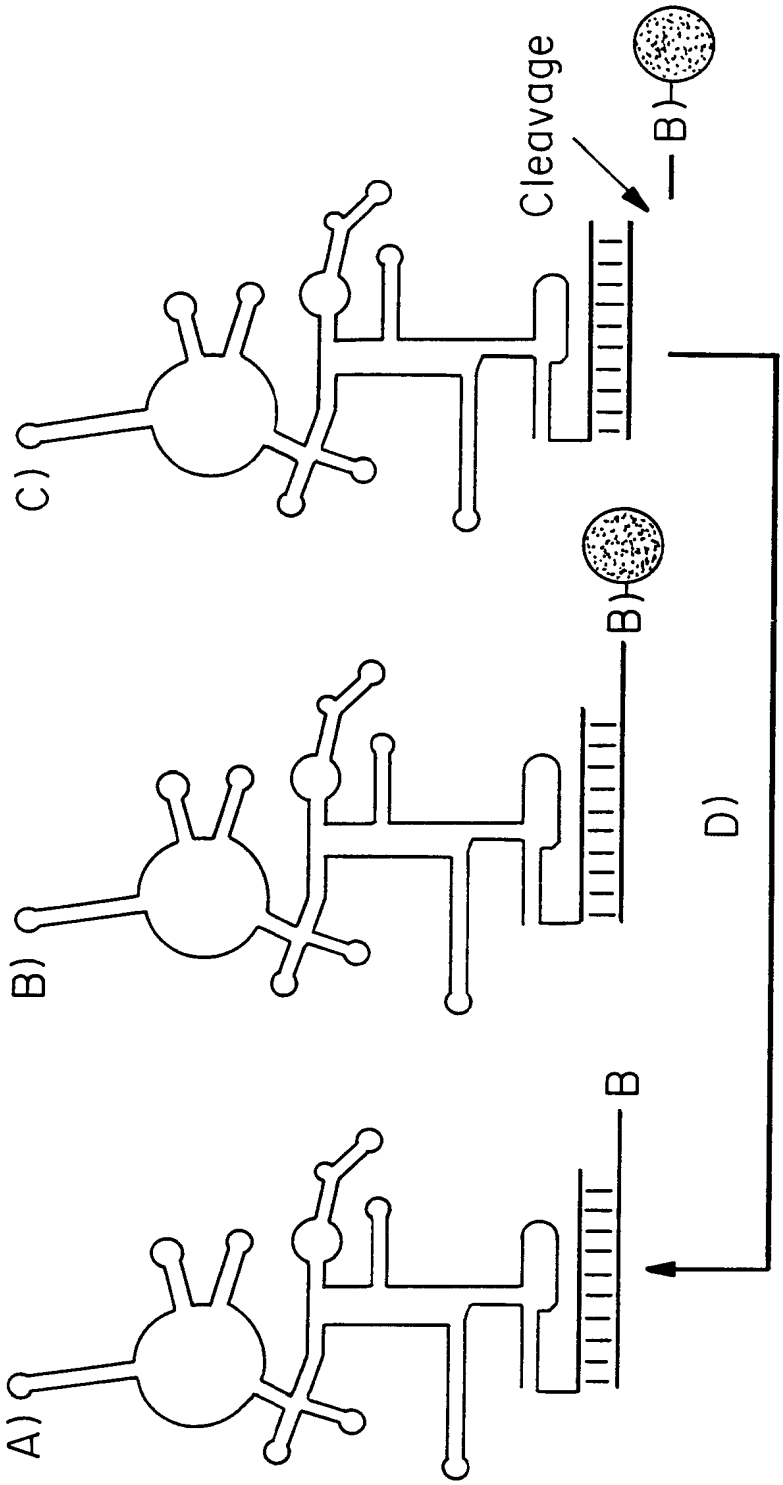


FIG. 8

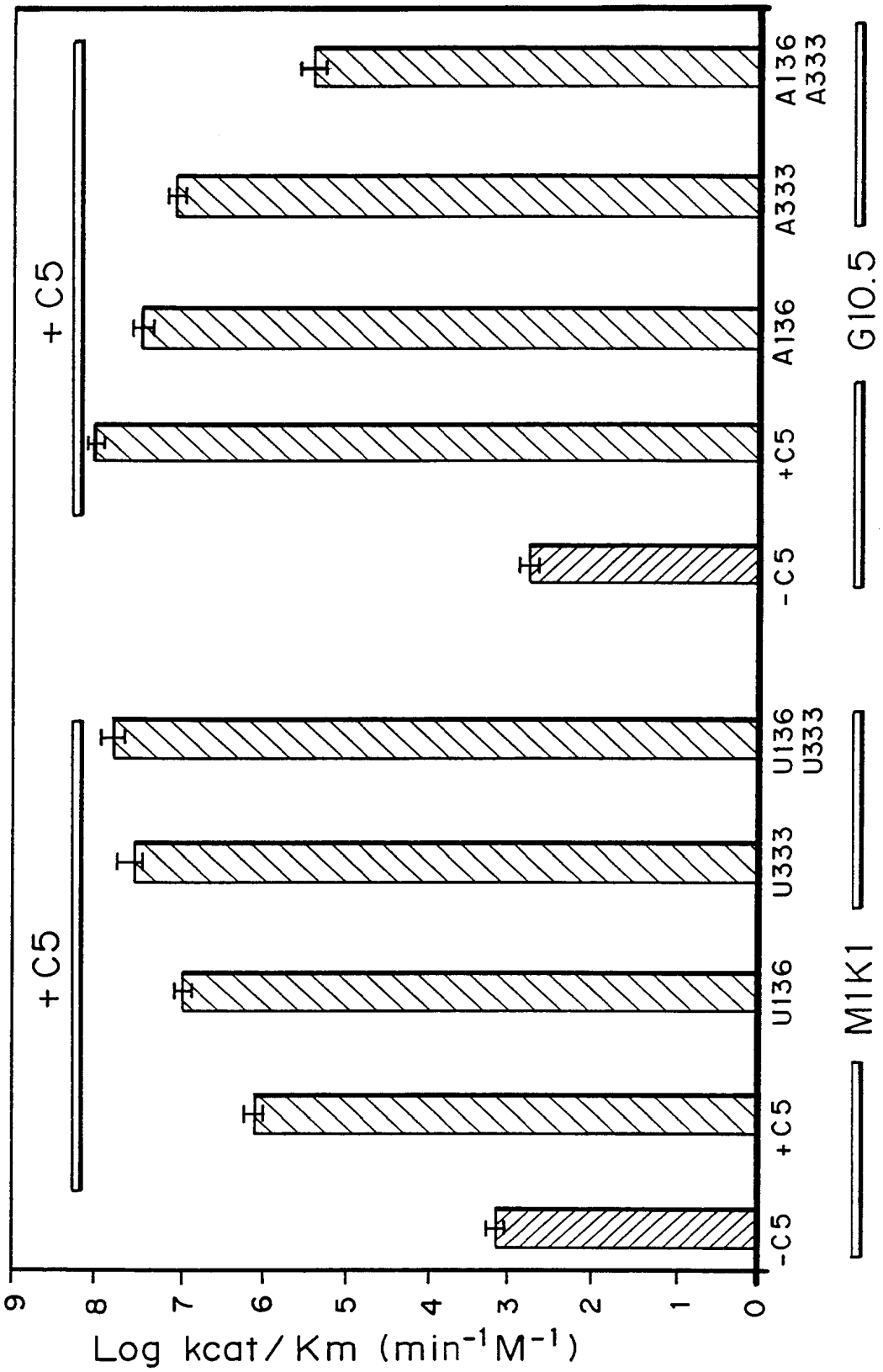

Log kcat/Km ($min^{-1}M^{-1}$)
0
1
2
3
4
5
6
7
8
9
-C5
+C5
U136
U333
+C5
U136
U333
-C5
+C5
A136
A333
+C5
A136
A333
M1K1
G10.5

FIG. 9

INTERNATIONAL SEARCH REPORT

Inte. onal Application No
PCT/US 00/26864

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/11 C12N9/00 C12N9/22 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, BIOSIS, MEDLINE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	LI YING ET AL: "Cleavage by RNase P of gene N mRNA reduces bacteriophage lambda burst size." NUCLEIC ACIDS RESEARCH, vol. 24, no. 5, 1996, pages 835-842, XP002160942 ISSN: 0305-1048 cited in the application figure 3; table 1 page 839, right-hand column, line 24 -page 840, left-hand column, line 6 --- -/--	1,23,35, 36

[X] Further documents are listed in the continuation of box C. [ ] Patent family members are listed in annex.

° Special categories of cited documents :

'A' document defining the general state of the art which is not considered to be of particular relevance

'E' earlier document but published on or after the international filing date

'L' document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

'O' document referring to an oral disclosure, use, exhibition or other means

'P' document published prior to the international filing date but later than the priority date claimed

'T' later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

'X' document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

'Y' document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

'&' document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
21 February 2001	06/03/2001

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Andres, S

1

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Inte onal Application No
PCT/US 00/26864

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
A	KIM JOSEPH J ET AL: "The protein cofactor allows the sequence of an RNase P ribozyme to diversify by maintaining the catalytically active structure of the enzyme." RNA (NEW YORK), vol. 3, no. 6, 1997, pages 613-623, XP000982393 ISSN: 1355-8382 cited in the application the whole document ---	10-22, 29-34, 40,41
A	CADWELL, C. & JOYCE, G.: "Mutagenic PCR" PCR METHODS & APPLICATIONS., vol. 3, 1994, pages s136-s140, XP002160943 ISSN: 1054-9803 cited in the application the whole document ---	1-42
A	LUMELSKY N ET AL: "SELECTION AND CHARACTERIZATION OF RANDOMLY PRODUCED MUTANTS IN THE GENE CODING FOR M1 RNA" JOURNAL OF MOLECULAR BIOLOGY, vol. 202, no. 3, 1988, pages 443-454, XP000982401 ISSN: 0022-2836 page 448; figure 4 page 450; table 1 page 451; table 4 ---	2,24
A	FRANK DANIEL N ET AL: "In vitro selection for altered divalent metal specificity in the RNase P RNA." PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES, vol. 94, no. 26, 23 December 1997 (1997-12-23), pages 14355-14360, XP002160944 ISSN: 0027-8424 cited in the application figure 3 ---	2,3,24, 25
A	TSANG JOYCE ET AL: "Specialization of the DNA-cleaving activity of a group I ribozyme through in vitro evolution." JOURNAL OF MOLECULAR BIOLOGY, vol. 262, no. 1, 1996, pages 31-42, XP002160945 ISSN: 0022-2836 --- -/--	

1

INTERNATIONAL SEARCH REPORT

Inter ›nal Application No
PCT/US 00/26864

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
P,X	COLE KYLE B ET AL: "Acquisition of novel catalytic activity by the M1 RNA ribozyme: The cost of molecular adaptation." JOURNAL OF MOLECULAR BIOLOGY, vol. 292, no. 4, 1 October 1999 (1999-10-01), pages 931-944, XP002160946 ISSN: 0022-2836 cited in the application the whole document	1-42

1